Functional characterization of

the FET family of RNA-binding proteins

Dissertation

zur Erlangung des akademischen Grades

doctor rerum naturalium

(Dr. rer. nat.)

im Fach Biologie

eingereicht an der Mathematisch-Naturwissenschaftlichen Fakultät I

der Humboldt-Universität zu Berlin

von

Diplom-Biologin Kerstin Baethge

Präsident der Humboldt-Universität zu Berlin

Prof. Dr. Jan-Hendrik Olbertz

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät I

Prof. Dr. Stefan Hecht

Gutachter: 1. Prof. Dr. Christian Schmitz-Linneweber

2. Dr. Markus Landthaler

3. Prof. Dr. Oliver Daumke

Tag der mündlichen Prüfung: 26.06.2014

For Steve, Lana, Levi and Niklas

SUMMARY

Post-transcriptional regulation of gene expression takes place at multiple levels between transcription and decay of the mRNA. RNA-binding proteins (RBPs) play a key role in orchestrating splicing, export, stability, localization and translation of mRNAs. FUS, EWSR1 and TAF15 constitute the FET protein family which participates in multiple levels of cellular function. FET proteins have been implicated to function in various cellular processes including transcription, pre-mRNA splicing and miRNA processing. Translocations and mutations in FET proteins lead to diverse pathologies. FUS is involved in neurodegenerative diseases like frontotemporal lobar degeneration (FTLD) and amyotrophic lateral sclerosis (ALS).

In this study, Photoactivatable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation (PAR-CLIP) was used to determine RNA-targets and binding sites of FUS, EWSR1 and TAF15, an ALS-causing FUS mutant and another ALS-related protein, TARDBP. The identified binding sites of FET proteins were mainly intronic, supporting the involvement of FUS and EWSR1 in splicing, which was validated by FET protein knockdown. Comparison of FUS and TARDBP RNA targets revealed that ubiquitin-proteasome related gene categories were overrepresented, further illustrating that aberrations in protein degradation are implicated in the pathogenesis of ALS. In addition, it was shown that FUS and TAF15 proteins preferentially bind UAC rich, single-stranded RNA sequences. mRNA sequencing after FUS, EWSR1 and TAF15 depletion in HEK293 cells revealed a stabilizing effect on their targets.

Interestingly, FET proteins also seem to influence transcription by interaction with promoter-associated noncoding RNAs.

In summary, we identified the RNA-targets and binding sites of all human FET proteins in comparison with an ALS-causing FUS mutant and TARDBP. Functional studies revealed an involvement of FET proteins in mRNA stabilization, splicing and transcriptional regulation.

Keywords: RNA-binding proteins, FUS, EWSR1, TAF15, TARDBP, PAR-CLIP, ALS

ZUSAMMENFASSUNG

RNA-bindende Proteine (RBPs) spielen eine zentrale Rolle in der posttranskriptionellen Kontrolle von mRNAs, die zwischen Transkription und Abbau von mRNAs stattfindet. RBPs beeinflussen Spleißen, Export, Stabilität, Lokalisierung und Translation von mRNAs. FUS, EWSR1 und TAF15 gehören zu der Familie der FET Proteine. Diese wirken an verschiedenen zellulären Prozessen wie Transkription, Spleißen und der Prozessierung von miRNAs mit. Translokationen und Mutationen der FET Proteine führen zu verschiedenen Krankheiten. FUS spielt eine Rolle bei den neurodegenerativen Krankheiten frontotemporale Lobärdegeneration (FTLD) und amyotrophe Lateralsklerose (ALS).

In dieser Arbeit wurde die mithilfe von photoaktivierbaren Ribonukleotiden UV-Licht induzierte Quervernetzung und Immunpräzipitation (PAR-CLIP) Methode genutzt, um die RNA-Bindestellen von FUS, EWSR1 und TAF15, einer ALS-verursachenden FUS Mutante und einem anderen, mit ALS in Verbindung stehenden Protein, TARDBP, zu bestimmen. Die RNA-Bindestellen der FET-Proteine lagen größtenteils in Introns. Passend dazu konnte durch knockdown der FET Proteine eine Rolle von FUS und EWSR1 im Spleißen von mRNAs validiert werden. Dem Ubiquitin-Proteasom-System zugehörige RNAs waren unter den sowohl von FUS als auch TARDBP gebundenen mRNAs überrepräsentiert. Dies bestätigt die Annahme, dass Störungen in der Proteindegradation die ALS-Pathogenese beeinflussen. Zusätzlich konnte gezeigt werden, dass FUS und TAF15 bevorzugt UAC-reiche, einzelsträngige RNA-Sequenzen binden. Sequenzierung von mRNAs nach Depletion von FUS, EWSR1 und TAF15 in HEK293-Zellen zeigte einen stabilisierenden Effekt der FET-Proteine auf gebundene mRNAs. Desweiteren scheinen die FET Proteine durch Interaktion mit Promotor-assoziierten, nicht-kodierenden RNAs die Transkription zu beeinflussen.

Schlagworte: RNA-Bindeproteine, FUS, EWSR1, TAF15, TARDBP, PAR-CLIP, ALS

Table of contents

List of figures

List of tables

1. Introduction

1.1 Posttranscriptional gene regulation by RNA-binding proteins

In eukaryotes, transcription and translation are locally separated in the nucleus and cytoplasm. Therefore, pre-mRNAs can undergo extensive post-transcriptional processing to achieve more diversity and introduce additional layers of gene regulation. Pre-mRNA processing includes splicing, capping, polyadenylation and editing of protein-coding transcripts. Consequently, post-transcriptional regulation of gene expression takes place at multiple levels between transcription and decay. Besides small non-coding RNAs, RNA-binding proteins (RBPs) play a key role in post-transcriptional control of mRNAs by orchestrating splicing, export, stability, localization and translation of mRNAs (Fig.1).

The human genome encodes around 600 proteins with RNA-binding domains (de Lima Morais et al. 2011; Ray et al. 2013), whereas the number of known RNA-binding domains is relatively small. Experimentally, two recent studies expanded this set to about 1100 mRNA-binding proteins (Baltz et al. 2012; Castello et al. 2012). Often RBPs contain more than one RNA-binding domain which in combination ensure specificity and affinity of binding (Lunde et al. 2007). In addition, several domains allow protein-protein interactions to form ribonucleoprotein particles (RNPs). In combination with additional functional domains the diversity of RBPs increases and allows a single RBP to bind multiple targets.

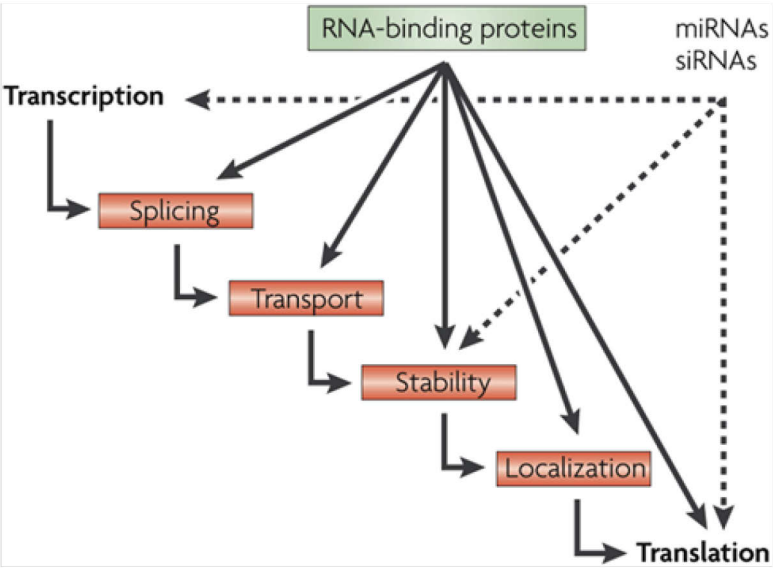


Fig. 1: Posttranscriptional gene regulation.
In eukaryotic cells, mRNAs undergo several steps of regulation from transcription to translation. The coordination of multiple mRNAs is regulated by RNA-binding proteins and small non-coding RNAs at different levels. miRNAs, microRNAs; siRNAs, small interfering RNAs. Modified from (Keene 2007).

1.2 Structure and functions of RNA-binding proteins

RBPs are often composed of multiple copies of a just few functional domains. Combination of different domains creates versatility in RNA binding with high affinity and high specificity. This modular structure of RBPs gives rise to a large functional repertoire of these proteins (Burd and Dreyfuss 1994). Further advantages of the modular protein structure are that these proteins can bind longer stretches of nucleic acids, sequences which are separated by nucleotides and sequences from different RNA molecules. This section will briefly summarize the most common RNA-binding domains and functions (Fig.2).

1.2.1 The RNA-recognition motif

One of the most abundant and best characterized RNA-binding domains in eukaryotes is the RNA-recognition motif (RRM). It is composed of 80-90 amino acids that form a four-stranded ß-sheet against two α-helices (Oubridge et al. 1994). RNA recognition usually occurs on the surface of the ß-sheet. A single RRM can recognize two (Mazza et al. 2002) to eight nucleotides (Price et al. 1998) but multiple domains are often needed to define sequence specificity because the number of recognized nucleotides by a single RRM is often too small to define a unique binding sequence (Auweter et al. 2006). RRM-domain-containing proteins are involved in many cellular functions, for instance mRNA and ribosomal RNA processing, splicing, translation, RNA export and stability (Dreyfuss et al. 2002). One example is the poly(A)-binding protein which regulates translation initiation (Kahvejian et al. 2005).

1.2.2 The K-homology domain

The K-homology domain (KH domain) was initially identified as a repeated sequence in the heterogenous nuclear ribonucleoprotein (hnRNP) K. The KH domain is composed of an evolutionary conserved sequence of around 70 amino acids which form a three stranded ß-sheet packed against three α-helices (Grishin 2001). It can bind single-stranded (ss) DNA as well as ssRNA (Backe et al. 2005) recognizing four nucleotide long sequences.

KH domains are found in RBPs with different functions including splicing, transcriptional regulation and translational control like the neuronal splicing factor Nova-1 (Lewis et al. 1999).

1.2.3 The Zinc finger domain

Zinc finger (ZF) were initially described as DNA-binding domains (Miller et al. 1985) but more recently, it was shown that ZF can also bind to RNA, protein and lipids (Matthews and Sunde 2002; Hall 2005; Gamsjaeger et al. 2007). A classical ZF domain is around 30 amino acids long which form a ββα structure held together by a Zn^{2+} ion. They are further classified depending on the amino acids that are interacting with this ion and are generally present in multiple repeats per protein. The different classes of ZFs differ largely in function as they mediate the interaction of proteins with other biomolecules. These classes have a variety of different roles within the cell like transcriptional regulation, mRNA processing and ubiquitination (Aasland et al. 1995; Lu et al. 2002; Loughlin et al. 2009).

1.2.4 The double-stranded RNA-binding motif

In contrast to the RNA-binding domains described above, double-stranded RNA-binding motifs (dsRBMs) were first described to recognize RNA structure rather than RNA sequence (Stefl et al. 2005). They are 70-90 amino acid long sequences which exhibit a conserved αββα protein topology. DsRBMs are often found in multiple repeats and are involved in a variety of functions within the cell like RNP localization, RNA interference, RNA processing, RNA localisation, RNA editing and translational control (Chang and Ramos 2005). Recently, the structure of ADAR dsRBMs in complex with RNA also revealed sequence specificity in dsRNA-binding (Stefl et al. 2010).

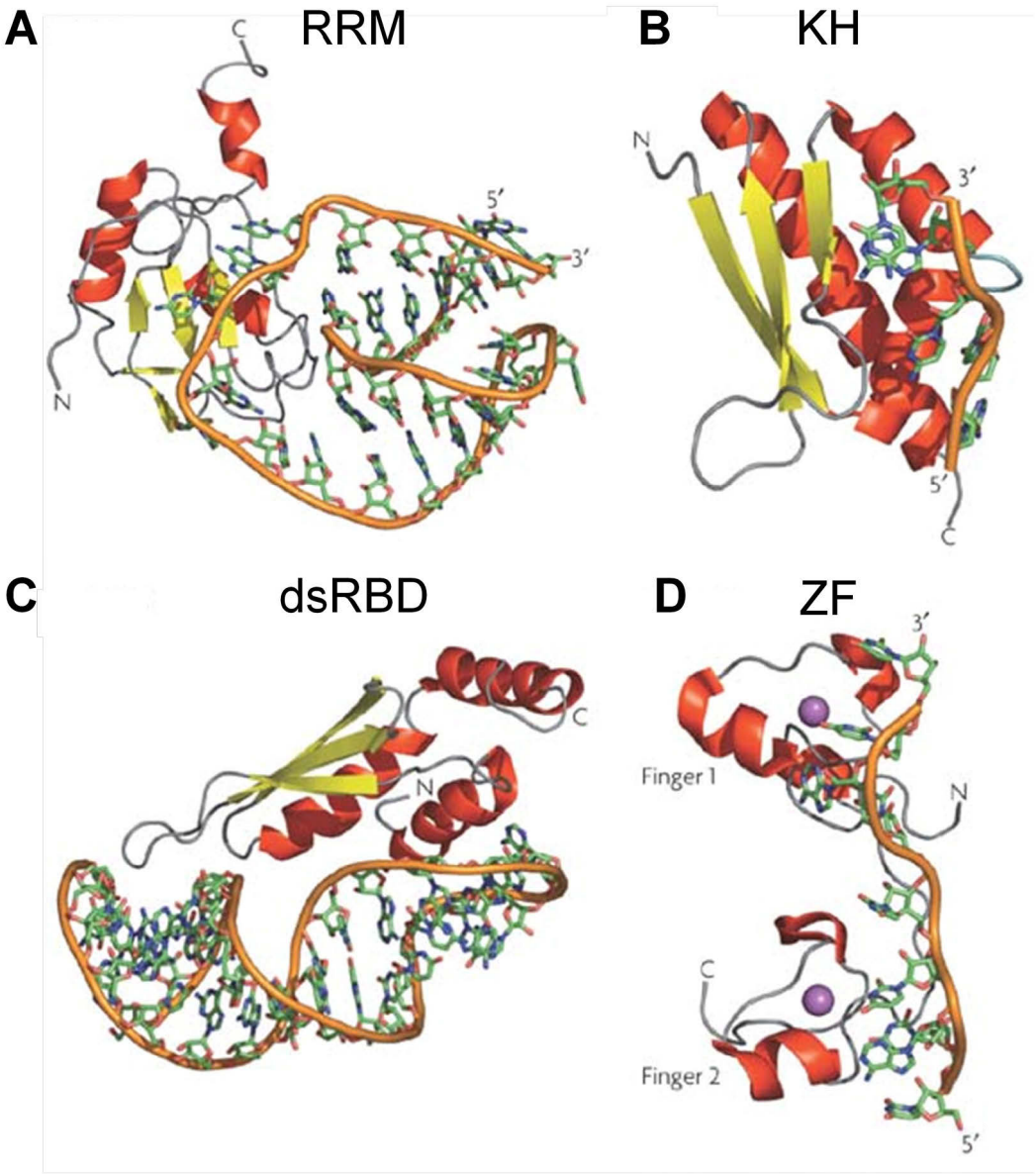


Fig. 2: Structures of RNA-binding domains in complex with RNA.

(**A**) Structure of the N-terminal RRM of human U1A bound to RNA. (**B**) The KH3 domain of Nova-2 bound to 5'-AUCAC-3'. (**C**) The yeast Rnt1 dsRBD bound to an RNA helix capped by an AGNN tetraloop. (**D**) The two zinc fingers of TIS11d bound to an AU-rich RNA element. In all panels, the RNA backbone is represented with an orange ribbon, α-helices are in red and β-sheets are in yellow; the zinc atom in the TIS11d structure is in magenta. Modified from (Lunde et al. 2007).

1.3 The FET protein family

Fused in sarcoma (FUS), Ewing sarcoma breakpoint region 1 (EWSR1) and TATA-binding protein-associated factor 15 (TAF15) constitute a family of RNA/DNA-binding proteins, known as FET proteins (Bertolotti et al. 1996). FET proteins are highly expressed in almost all human fetal and adult tissues and are predominantly located in the nucleus, although they are able to shuttle between the nucleus and cytoplasm (Andersson et al. 2008).

1.3.1 Structure and binding preferences of the FET proteins

The FET proteins are structurally very similar and composed of several domains (Fig.3). The N-terminal end contains a serine-tyrosine-glycine-glutamine-rich domain (SYGQ-domain), which can serve as transcriptional activation domain when fused to other transcription factors in cancer associated translocations (Zinszner et al. 1994; Bertolotti et al. 1999). At their C-termini, FET proteins harbour domains that are implicated in nucleic acid binding. The most conserved region is an RRM (Tan and Manley 2009), which is flanked by several arginine-glycine-glycine-rich domains (RGG) and a RanBP2-type ZF. RanBP2-type ZFs were shown to bind single-stranded RNA with preference to GGU-containing motifs (Loughlin et al. 2009; Nguyen et al. 2011). FET proteins have been reported to bind both RNA as well as ss and dsDNA (Bertolotti et al. 1996; Hackl and Luhrmann 1996). Recently, a study by Takahama et al. showed that the RGG-rich domain of FUS forms a ternary complex with the human telomere G-quadruplex DNA and telomeric repeat-containing RNA in vitro (Takahama et al. 2013). This observation indicates that the RGG-rich domain might be also responsible for the DNA-binding properties of the FET proteins. The region with the highest degree of sequence identity between all three FET family members is the RRM (Tan and Manley 2009). It folds into a secondary structure which might be involved in sequence specific RNA binding (Burd and Dreyfuss 1994; Hackl and Luhrmann 1996). Together with the RGG-rich domains, these regions are essential for the RNA-binding specificity of the FET proteins. Sequence-specific binding was reported for both FUS and EWSR1, indicating that the proteins interact with polyU and polyG stretches (Ohno et al. 1994). Furthermore, it was shown that FUS binds specifically to a GGUG motif (Lerga et al. 2001) and AU-rich stem-loop structures (Hoell et al. 2011). The RNA-binding specificity of TAF15 remains elusive.

At their very C-terminal end all three FET proteins harbour a nuclear localization signal (NLS) which mediates transport of the proteins into the nucleus (Zakaryan and Gehring 2006; Dormann et al. 2010; Marko et al. 2012).

Nuclear import of the FET proteins is mediated by Transportin 1 and 2 (Dormann et al. 2010) whereas nuclear import of EWSR1 is also dependent on the phosphorylation state of the C-terminus (Leemann-Zakaryan et al. 2011). Taken together, the domain composition suggests diverse functions of the FET proteins (see next section).

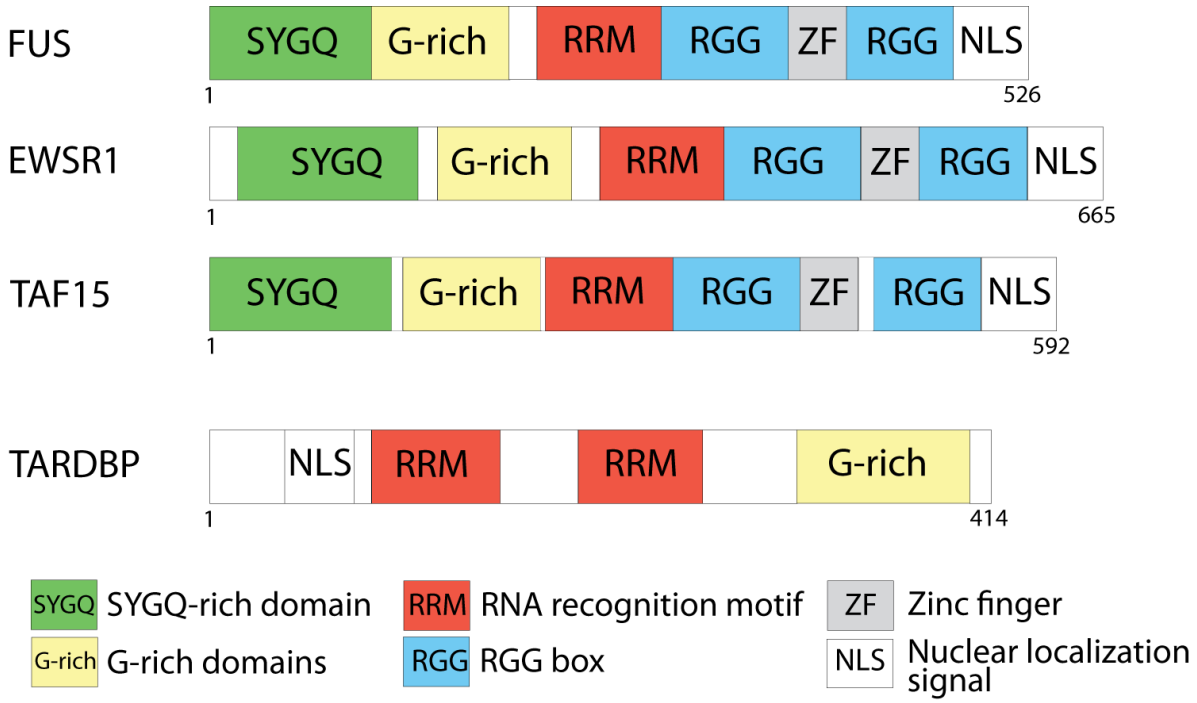


Fig. 3: Domain structure of the FET protein family and TARDBP.
FUS, EWSR1 and TAF15 share the same domain structure. The N-terminal end consists of a SYGQ-rich domain. The C-terminus is composed of a G-rich domain, a RRM domain, a ZF domain flanked by two RGG boxes and a nuclear localization signal. The unrelated TARDBP consists of an N-terminal nuclear localizations signal, two RRM domains and a C-terminal G-rich domain.

1.3.2 Functions of the FET proteins

The FET proteins seem to be implicated in various cellular processes. The next section will give an overview of the current state of knowledge about functions of FUS, EWSR1 and TAF15.

1.3.2.1 Transcriptional regulation

Several lines of evidence suggest an involvement of FET proteins in transcription. FUS, EWSR1 and TAF15 were found to interact with distinct subpopulations of the RNA Polymerase II (RNA Pol II) associated complex TFIID as well as RNA Pol II subunit hRPB3 (Bertolotti et al. 1996; Hoffmann and Roeder 1996; Bertolotti et al. 1998). Transcription factor II D (TFIID) is composed of the TATA box binding protein (TBP) and TBP-associated factors (TAF(II)s) and is involved in initiation and elongation of transcription (Workman and Roeder 1987). In addition, FUS and EWSR1 were reported to interact with various transcription factors like POU4F1 (Thomas and Latchman 2002) and OCT4 (Lee et al. 2005). These observations suggest an involvement of FUS and EWSR1 in transcriptional regulation. Moreover, FET proteins function in a so far unique case of transcriptional regulation. FUS, EWSR1 and TAF15 bind to non-coding RNAs (ncRNA) derived from the promoter of the cyclin D1 gene upon DNA damage, which is transcriptionally repressed *in cis* through inhibition of the p300 histone acetyltransferase by a FET-ncRNA ribonucleoprotein complex (Wang et al. 2008). Besides their role in RNA Pol II mediated transcription, recent findings also suggest FUS repressing RNA Pol III transcription of small untranslated RNAs (Tan and Manley 2010).

1.3.2.2 Splicing

Many studies associated FET proteins with pre-mRNA splicing as they were identified as part of the spliceosome (Rappsilber et al. 2002). Originally, FUS was identified as the hnRNP P2 belonging to a group of proteins involved in pre-mRNA-processing (Calvio et al. 1995). Furthermore, FUS and EWSR1 are interacting with multiple splicing factors like YBX1 (Chansky et al. 2001) and serine-arginine proteins such as TASR and SC35 (Yang et al. 1998; Yang et al. 2000). In addition, FUS associates with hnRNP A1 and C1/C2 as well as with SRm160 and PTB (Lerga et al. 2001; Meissner et al. 2003). Moreover, FUS was found to bind to the pre-mRNA 3' splice site and seemed to promote the usage of distal 5' splice sites (Wu and Green 1997; Hallier et al. 1998). The interaction of FUS with the splicing machinery appears to have functional consequences since splicing of pre-mRNAs expressed from several minigenes is affected by the FET proteins (Hallier et al. 1998; Chansky et al. 2001; Kino et al. 2011).
Recently, two studies showed that FUS has an effect on alternative splicing with increased binding of FUS to introns around repressed exons using brain from FUS -/- mice and *FUS*-silenced primary cortical neurons, respectively (Ishigaki et al. 2012; Rogelj et al. 2012).

Similarly, Paronetto and coworkers could show that depletion of EWSR1 in HeLa cells leads to changes in alternative splicing of DNA damage-induced genes (Paronetto et al. 2011). It would be conceivable that the FET proteins co-regulate transcription and splicing as they are able to bind both RNA Pol II and various splicing factors and since transcription is physically and functionally coupled to splicing (Montes et al. 2012).

1.3.2.3 mRNA transport

Since the FET proteins can shuttle between the nucleus and cytoplasm (Zinszner et al. 1997) they have been also implicated in mRNA transport. FUS is localized in dendrites of mouse hippocampal neurons and is transported to spines upon activation of the glutamate receptor 5 (Fujii et al. 2005). One mRNA that is transported by FUS to dendritic spines is Nd1-L, encoding an actin-stabilizing protein which may play a role in the dynamic organization of the actin cytoskeleton (Sasagawa et al. 2002; Fujii and Takumi 2005). Whether EWSR1 and TAF15 also play a role in RNA transport and cytoskeleton stabilization has not been investigated yet. FUS and TAF15 are also implicated in regulation of localized protein synthesis since they are accumulating together with other RBPs in spreading initiation centers of adhering cells (de Hoog et al. 2004; Andersson et al. 2008).

1.3.2.4 miRNA biogenesis

Moreover, FET proteins seem to participate in the regulation of miRNA processing as they were identified as part of the large Drosha complex (Gregory et al. 2004). For FUS and EWSR1, direct interaction with Drosha was demonstrated recently (Morlando et al. 2012; Sohn et al. 2012). Drosha is a nuclear RNase III enzyme which processes pri-miRNAs to pre-miRNAs as part of the microprocessor complex. Besides Drosha, also DGCR8 belongs to the microprocessor complex (Lee et al. 2003). DGCR8 contains an RNA-binding domain and is thought to bind and recognize pri-miRNAs for processing by Drosha (Yeom et al. 2006). FUS was identified as one of DGCR8-associated proteins (Shiohama et al. 2007).
In addition, FUS regulates miRNA biogenesis by binding specific pri-miRNAs involved in neuronal function and differentiation (Morlando et al. 2012). EWSR1 might directly or indirectly function in the maturation of let-7g as depletion of EWSR1 lead to an accumulation of pre-let-7g but downregulation of mature let-7g (Sohn et al. 2012).

A recent study by Ballarino and colleagues could show that also TAF15 is involved in miRNA mediated regulation of *CDKN1A* expression which is a key regulator of cell cycle and cell death (Ballarino et al. 2012). Hence, the FET proteins might also be important regulators of miRNA maturation but the exact mechanism remains elusive.

1.3.2.5 Genome surveillance

Additionally, the FET proteins have been associated with genomic surveillance and DNA repair. FUS knockout mice show male sterility and high genomic instability (Kuroda et al. 2000). FUS knockout mice as well as EWSR1 deficient mice have a defective B-cell development and show enhanced sensitivity to radiation (Hicks et al. 2000; Li et al. 2007). Furthermore, inactivation of EWSR1 in embryonic fibroblasts resulted in reduced meiotic recombination and premature cellular senescence (Li et al. 2007). Together with the ability of all three FET proteins to mediate pairing of homologous DNA ends, this suggests a role in DNA repair (Baechtold et al. 1999; Bertrand et al. 1999; Guipaud et al. 2006). This hypothesis is strengthened by the interaction of EWSR1 with the BRCA1-associated ring finger domain protein BARD1 (Spahn et al. 2002). BARD1 binds to the breast cancer susceptibility gene *BRCA1* that provides a platform for interactions with proteins involved in DNA repair and checkpoint control (Venkitaraman 2001).

Taken together, the FET family are involved in multiple cellular functions like transcription, splicing, RNA transport, miRNA biogenesis and DNA repair suggesting a possible role as master regulators in the cell.

1.3.3 FET proteins and their involvement in diseases

Recent studies revealed an association of FET proteins with neurological disorders. Mutations in the C-terminus of FUS can lead to amyotrophic lateral sclerosis (ALS), a fatal neurodegenerative disease characterized by a late-onset premature loss of upper and lower motor neurons in the cerebral cortex, brainstem and spinal cord (Kwiatkowski et al. 2009; Vance et al. 2009). This leads to a progressive skeletal muscle atrophy, causing death within 2 to 5 years due to respiratory failure (Kiernan et al. 2011). About 10% of all ALS cases are dominantly inherited, whereas the remaining cases are sporadic.

Besides mutations in other genes like superoxide dismutase 1 (*SOD1*) and *C9orf72* about 4% of all familial ALS cases and rare sporadic cases are caused by mutations in the *FUS* gene. The mutant FUS protein is mislocalized to the cytoplasm accumulating in ubiquitin-positive inclusion bodies in neurons and glial cells of brain and spinal cord of ALS patients (Kwiatkowski et al. 2009). Most of the FUS mutations causing ALS are located at the C-terminus where the NLS is located, leading to disrupted Transportin binding and disturbed nuclear import of the protein (Dormann et al. 2010; Ito et al. 2011). The mutated protein is recruited into stress granules (Bosco et al. 2010), cytosolic structures composed of temporally stored mRNAs and associated RBPs, which form upon environmental stresses like oxidative stress or heat shock (Anderson and Kedersha 2008).

A study by Daigle and coworkers showed that incorporation into stress granules of the FUS mutant protein is dependent on the RNA-binding ability of the protein. Furthermore, RNA-binding of FUS regulates cytoplasmic mislocalization and neurodegeneration since RNA-binding-incompetent FUS mutants block the neurodegenerative phenotype in a Drosophila ALS model and neuronal cell line (Daigle et al. 2013). In contrast, a recent study by Shelkovnikova and colleagues revealed that FUS aggregation is sufficient to cause an ALS-like phenotype in transgenic mice (Shelkovnikova et al. 2013). Expression of a FUS variant lacking the RNA-binding domain and the NLS in transgenic mice causes severe damage of motor neurons suggesting that aggregation of FUS protein can by itself trigger neuroinflammation independent of its roles in RNA metabolism.

Recent reports also implicate EWSR1 and TAF15 mutations in ALS (Couthouis et al. 2011; Ticozzi et al. 2011; Couthouis et al. 2012). Similar to ALS, FUS-containing inclusion bodies were also found in sub-population of patients with frontotemporal lobar degeneration (FTLD), which is characterized by degeneration of frontal and temporal cortical neurons, confirming the long-standing thought that the two diseases are related (Neumann et al. 2009).

Nevertheless, the pathology of FTLD-FUS is slightly different since all three FET proteins co-localize in pathological inclusion bodies whereas in ALS-FUS patients no co-deposition of EWSR1 and TAF15 in FUS-positive inclusions was observed (Neumann et al. 2011). Interestingly, mutations in a second RNA/DNA-binding protein, TARDBP/TDP-43, were also discovered in patients with ALS and FTLD. TARDBP is structurally unrelated to FET proteins (Fig. 3) and likewise ubiquitin-positive cytoplasmic inclusions containing TARDBP are observed in disease-affected tissues (Neumann et al. 2006).

In contrast to mutations in FUS, most of the identified ALS causing mutations in TARDBP are clustered in the C-terminal glycine-rich region which mediates interaction with other hnRNP proteins regulating splicing of pre-mRNAs (Buratti et al. 2005). Unlike FUS, none of the identified mutations so far are located in the NLS of TARDBP indicating no effect on nuclear transport of the protein.

Interestingly, FUS and TARDBP were found in direct interaction as part of a biochemical complex regulating histone deacetylase 6 (HDAC6) mRNA levels (Kim et al. 2010). ALS-causing mutations of TARDBP also increase stability of the mutant protein and promote complexes with FUS (Ling et al. 2010). These genetic findings and the common pathology indicate that FUS and TARDBP proteins abnormally aggregate in ALS and FTLD, and suggest similar molecular mechanisms aberrantly regulated at the post-transcriptional level as potential pathogenic clues, although the primary or secondary role of each of these events in triggering motor neuron degeneration still need to be determined.

In addition, genetic aberrations in FET proteins are also associated with several other human diseases. FUS and EWSR1 were initially discovered to be chromosomally translocated in sarcomas (Delattre et al. 1992; Crozat et al. 1993), which are aggressive cancers of the supportive and connective tissue in the human body. Chromosomal translocation results in the fusion of the N-terminal transcriptional activator domain of the FET proteins to the DNA binding domain of various transcription factors. This leads to aberrant transcriptional activation under the control of the FET protein promoter. *FUS* and *EWSR1* were both found to be fused to the transcription factors *CHOP* and *ERG*, for example, leading to the development of myxoid liposarcoma and Ewing's sarcoma family of tumours, respectively (Crozat et al. 1993; Rabbitts et al. 1993; Zucman et al. 1993). Furthermore, cancer associated fusion of *TAF15* to *ZNF384* has been observed in acute leukaemia (Martini et al. 2002). Translocation affects only one allele therefore tumour cells express both full-length FET proteins and fusion protein. It was shown that both protein forms are strongly expressed in tumour cells (Spitzer et al. 2011), so it is rather a gain-of-function of the fusion protein which leads to deregulated target expression and an altered differentiation pattern of the cells (Martini et al. 2002).

1.4 Outline and objectives of the thesis

RBPs play a key role in post-transcriptional control of mRNAs. Post-transcriptional regulation of gene expression takes place at multiple levels during the lifecycle of an mRNA. RBPs orchestrate splicing, export, editing, stability, localization and translation of mRNAs.
FUS, EWSR1 and TAF15 constitute the FET family of proteins which are involved in several steps of post-transcriptional regulation. In this thesis the diverse functions of all three members of the FET family were systematically studied in HEK293 cells as a model cell line.
Principal aim of this work was to characterize the regulatory functions and mechanisms on the transcriptional and post-transcriptional level of the FET family by using several systematic high-throughput approaches (Fig.4):

- A prerequisite for understanding the function of RBPs is a comprehensive identification of RBP binding sites. Therefore, PAR-CLIP methodology was used to define the target transcripts, binding sites and possible binding motifs of FUS, EWSR1 and TAF15 in comparison with an ALS causing FUS truncation mutant (FUS R495X) and TARDBP, another protein involved in ALS.
- Subsequently, the next aim was to identify at which step of gene expression these targets are regulated by siRNA mediated knockdown of all three FET proteins in HEK293 cells. Correlation of changes after FET protein depletion in
    - abundance and splicing of mRNAs was determined by mRNA sequencing (mRNA-Seq)
    - miRNA expression was examined by small RNA sequencing (small RNA-Seq)
    - protein abundance was analyzed using stable isotope labeling with amino acids in cell culture (SILAC)
    - transcription was investigated by chromatin immunoprepecipitation followed by quantitative polymerase chain reaction (ChIP-qPCR)

By relating the protein-RNA interaction maps to RNA sequencing data and the proteomic analysis of FET-depleted cells, functions and regulatory mechanisms of each FET family member on the post-transcriptional level can be deduced. Comprehensive protein-RNA interaction maps of the FET proteins are crucial to identify common or non-redundant regulatory functions.

By using several systematic high-throughput approaches for all three FET family members and TARDBP together with a disease related FUS mutant in one cell line it is feasible to correlate the genomic, transcriptomic and proteomic data sets.

Previous studies only indentified the RNA targets of FUS and TARDBP individually but not in comparison with EWSR1 and TAF15 (Polymenidou et al. 2011; Tollervey et al. 2011; Lagier-Tourenne et al. 2012; Rogelj et al. 2012). Only one recent study defined the global RNA targets of all three FET proteins in comparison with two ALS-causing FUS mutants (Hoell et al. 2011). Despite several biochemical studies investigating the function of FET proteins in various nuclear processes, the impact of FET proteins on RNA binding with respect to the development of ALS have been unexplained. Comparison of FET protein data with TARDBP and the FUS mutant data sets will likely shed more light into reasons and development for ALS which could be based on differences in RNA binding.

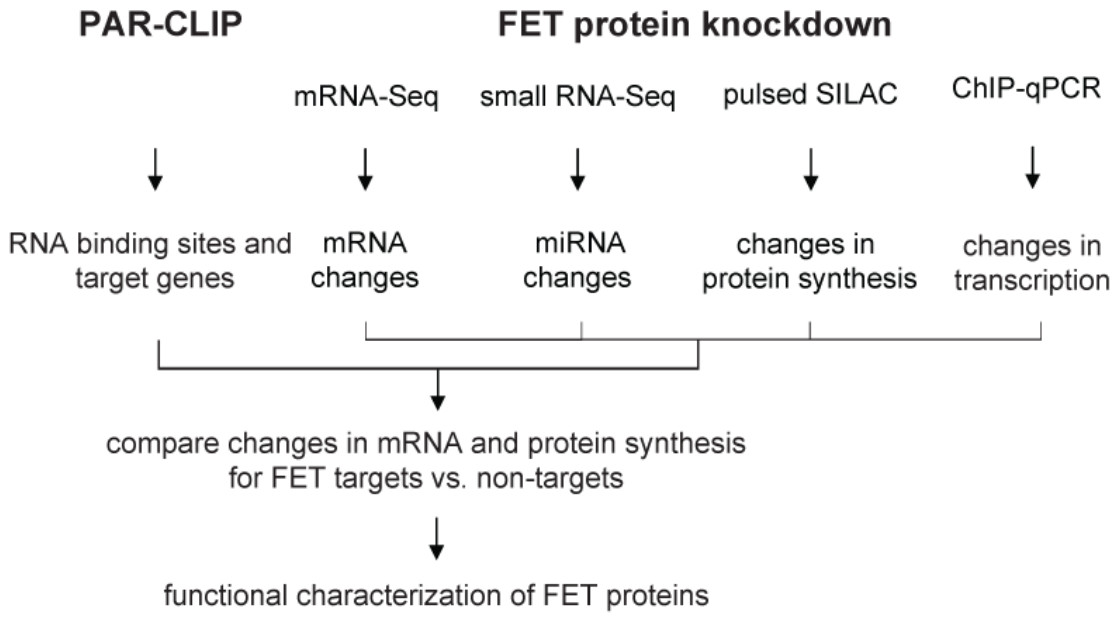


Fig. 4: Overview of the project.

Outline of the overall experimental approach. PAR-CLIP of FUS, EWSR1 and TAF15 were done in HEK293 cells. Together with siRNA mediated knockdown of the FET proteins followed by mRNA sequencing, small RNA sequencing, mass spectrometry and chromatin immunoprecipitation (ChIP) followed by qPCR changes in relative RNA and protein abundance of FET target mRNAs were elucidated.

2. Materials and methods

2.1 Cell lines and culture conditions

2.1.1 Cloning

Plasmids pENTR4 FUS, FUSR495X, EWSR1, TAF15 and TARDBP and were generated by polymerase chain reaction (PCR) amplification of the respective coding sequences (Tab.1) using HEK293 genomic DNA as a template (primers listed in 2.13.1). PCR was followed by restriction digest with SalI and NotI and ligation into pENTR4 (Invitrogen, UK). pENTR4 FUS, FUS R495X, EWSR1, TAF15 and TARDBP were recombined into pFRT/TO/FLAG/HA-DEST destination vector (Invitrogen, UK) using GATEWAY LR recombinase (Invitrogen,UK) according to the manufacturer's protocol to allow for doxycycline-inducible expression of stably transfected FLAG/HA-tagged protein in Flp-In T-REx HEK293 cells (Invitrogen, UK).

Tab. 1: NCBI Reference Sequence numbers of cloned coding sequences

Gene	NCBI Reference Sequence
FUS	NM_004960.3
EWSR1	NM_013986.3
TAF15	NM_139215.2
TARDBP	NM_007375.3

2.1.2 Cell lines and culture conditions

HEK293 T-REx Flp-In cells (Life Technologies, UK) were cultivated at 37°C and 5% CO_2 in DMEM high glucose (Life Technologies) with 10% (v/v) fetal bovine serum (Life Technologies, UK), 1% (v/v) 2mM L-glutamine (Life Technologies, UK), 1% (v/v) 10,000 U/ml penicillin/10,000 µg/ml streptomycin (Life Technologies, UK), 100 µg/ml zeocin (Invivogen, USA) and 15 µg/ml blasticidin (Invivogen, USA). SILAC medium was prepared as described previously (Ong & Mann, 2006).

Briefly, DMEM Glutamax lacking arginine and lysine (PAA, Austria) was supplemented with 10% dialyzed FBS (Sigma-Aldrich, Germany) and 2 mM L-glutamine (PAA, Austria). Amino acids (84 mg/l $^{13}C_6{}^{15}N_4$ L-arginine plus 146 mg/l $^{13}C_6{}^{15}N_2$ L-lysine or 84 mg/l $^{13}C_6$-L-arginine plus 146 mg/l D_4-L-lysine) were added to obtain „heavy“ and „medium-heavy“ medium, respectively. The corresponding non-labeled amino acids were used to prepare non-labeled “light” medium. All amino acids were purchased from Sigma-Aldrich (Germany).
HEK293 T-REx Flp-In cells stably expressing FLAG/HA-tagged proteins were generated by co-transfection with a 1:9 ratio of pFRT/TO/FLAG/HA constructs with pOG44 (Life Technologies, UK) using Lipofectamine 2000 (Life Technologies, UK). Cells were selected by exchanging zeocin with 100 mg/ml hygromycin (Sigma-Aldrich, Germany).

2.2 siRNA transfection

For knockdown experiments HEK293 T-REx Flp-In cells were grown in light SILAC medium. siRNA transfections of cells were performed in 6-well format using Lipofectamine RNAiMAX (Life Technologies, UK) as described by the manufacturer. One day before transfection $5x10^5$ cells were seeded. Transfections were carried out in SILAC-DMEM supplemented with 2 mM L-glutamine as transfection medium and 150 pmol siRNA and 7.5 µl Lipofectamine RNAiMAX for each transfection. Control transfections (mock) contained only the transfection reagent. Cells were harvested four days after transfection. Knockdown efficiency of proteins was checked by Western blot analysis and quantitative real-time polymerase chain reaction (qRT-PCR).

2.3 RNA extraction

Total RNA was extracted using the miRNAeasy Mini kit (Qiagen, Germany) as described by the manufacturer. RNA concentration and quality was assessed using a NanoDrop ND-1000 UV-VIS Spectrophotometer (Thermo Fisher Scientific, USA).

2.4 Poly(A) RNA isolation

Poly(A) mRNA was purified from 1 µg of total RNA using the Dynabeads mRNA Purification Kit (Life Technologies, UK) according to the manufacturer's protocol. The eluate was hybridized to the same beads for the second extraction step. Depletion of ribosomal RNAs was validated by capillary gel electrophoresis on a Bioanalyzer (Agilent, USA). Poly(A) RNA was subsequently processed for sequencing (see below).

2.5 Transcriptome sequencing

The poly(A)+ mRNA fraction was used for the sequencing library preparation according to the NEBNext mRNA Sample Prep kit (NEB, USA) instructions, with modifications. The mRNA was eluted from the beads with 17 µl of 10 mM Tris-HCl (pH 7.5), combined with 4 µl of 5x fragmentation buffer, incubated for exactly 3.5 min at 94ºC and placed on ice. This procedure yields RNA fragments ranging from 60 to 200 nt. After fragmentation, the RNA was purified using Agencourt RNAClean XP beads (Beckman Coulter Genomics, USA) according to manufacturer's protocol. Complementary DNA (cDNA) synthesis, end repair, addition of A overhangs and ligation of the adapters were performed as described in the NEBNext mRNA Sample Prep kit, each step followed by purification on Agencourt AMPure XP beads (Beckman Coulter Genomics, USA). The library was then PCR-amplified using Phusion polymerase (Thermo Fisher Scientific, USA) for 15 cycles of 10 s at 98ºC, 30 s at 65ºC and 30 s at 72ºC. After purification on Agencourt AMPure XP beads (Beckman Coulter Genomics, USA), the concentration and quality of the library were assessed by gelelectrophoresis on the Bioanalyzer using the DNA 1000 kit (both Agilent Technologies, USA). dsDNA libraries subsequently processed for sequencing using the Genomic DNA Sample Prep Kit (Illumina, USA) according to the manufacturer's protocol. Libraries were sequenced on Illumina Genome Analyzer GAII or Illumina HiSeq (Illumina, USA) using the 2x76 bp paired-end protocol.

2.6 Small RNA sequencing

Small RNAs of knockdown cells were isolated from 10 µg total RNA using the FlashPage Gel system (Life Technologies, UK) and sequenced using the small RNA cloning protocol (Hafner et al., 2008) with barcoded pre-adenylated 3'adapters (2.13.3.2).

2.7 Quantitative Real-Time - Polymerase Chain Reaction (qRT-PCR)

Single stranded cDNAs were synthesized from total RNA with an oligo-d$(T)_{18}$ primer or random hexamer primers using Superscript III Reverse Transcriptase (Life Technologies, UK) according to the manufacturer's instructions.

RT-PCR was performed using Power SYBR Green PCR master mix (Applied Biosystem, USA) on the StepOne Real-Time PCR System (Applied Biosystem, USA) for 30 cycles of 15 s at 94°C, 15 s at 60°C, and 20 s at 72°C.

For quantification of miRNA levels, TaqMan Micro RNA Assays from Applied Biosystems (RNU24, RNU6B, hsa-miR-34a, has-miR-374b, hsa-miR-92a, hsa-miR-10a, has-miR-148a, has-miR-19a) and 2x TaqMan PCR Mastermix (Applied Biosystems, USA) were used according to manufacturer's instructions.

2.8 Quantification of alternative exon inclusion

Two µg of total RNA was reverse transcribed with an oligo-d$(T)_{18}$ primer using Superscript III Reverse Transcriptase (Life Technologies, UK) according to the manufacturer's instructions. PCR amplification was performed using the KOD Hot Start DNA Polymerase kit (Novagen, Germany), 0.3 µM of each of the forward and reverse primers, and 2µl of cDNA (10% of the reverse transcription reaction) for 27 - 30 cycles of 20 s at 95 ºC, 15 s at 60 ºC, and 15 s at 70ºC.

PCR products were purified using PCR purification kit (Qiagen, Germany) and resolved on 2 % TBE-agarose gels. In parallel, PCR products were analyzed by the BioAnalyzer DNA 1000 Assay (Agilent technologies, USA). PSI (percent spliced in) values were calculated as the molar ratio of the peak corresponding to the exon containing isoform and the sum of the peaks representing both isoforms.

2.9 Labeling of proteins, sample preparation and measurement by mass spectrometry

Protein extraction, sample preparation and analysis by liquid chromatography tandem mass spectrometry (LC-MS/MS) was conducted by Dr. Guido Mastrobuoni, member of the Integrative Proteomics and Metabolomics group at the Berlin Institute for Medical Systems Biology at the Max-Delbrück Center for Molecular Medicine (Robert-Rössle Str. 10, D-13125 Berlin, Germany)

Prior transfection cells were grown at least two weeks in light SILAC medium. Transfection was performed as described (see 2.2). 29 hours after transfection, siRNA and mock-transfected cells were transferred to medium-heavy and heavy SILAC medium, respectively. After 24 hours of labeling, cells were harvested and equal numbers of siRNA- and mock-transfected cells were pooled. Proteins were extracted, digested and analyzed by liquid chromatography tandem mass spectrometry (LC-MS/MS) on a high-resolution instrument (LTQ-Orbitrap Velos, Thermo Scientific, USA). Raw data were analyzed using the MaxQuant proteomics pipeline (v1.2.2.5).

2.10 Western blotting

Cells were collected and lysed in 3 volumes of NP40 lysis buffer (50 mM HEPES-K pH 7.5, 150 mM KCl, 2 mM EDTA, 0.5% (v/v) NP-40, 0.5 mM DTT, complete EDTA-free protease inhibitor cocktail (Roche, Switzerland)). Protein concentration was determined using Bradford reagent (Thermo Scientific, USA). Proteins were subjected to SDS-PAGE and transferred to nitrocellulose using semi-dry blotting apparatus (Bio-Rad, USA). Membranes were blocked for 2 h at RT with 5% (w/v) non-fat milk in TBST and incubated with primary antibodies at 4°C over night. HRP-conjugated secondary antibodies (Dako, Denmark) were incubated for 1 h at room temperature. Bands were visualized with Amersham ECL Western Blotting Detection Reagents (GE Healthcare, UK) on a Fujifilm LAS-4000 luminescent image analyzer (GE Healthcare, UK). Primary antibodies used were anti-FUS (Abcam, ab23439) 1:1000 diluted, anti-EWSR1 (Abcam, ab81971) 1:500 diluted, anti-TAF15 (Abcam, ab69581) 1:500 diluted, anti-TARDBP (Abcam, ab57105) 1:500 diluted, anti-HA (Covance, MMS-101P) 1:1000 diluted and anti-Tubulin (Sigma, T8328) 1:3000 diluted. Secondary antibodies used were anti-rabbit HRP (Dako, P0048) 1:2000 diluted, anti-mouse HRP (Dako, P0447) 1:1000 diluted and anti-goat HRP (Dako, P0449) 1:1000 diluted.

2.11 PAR-CLIP

The PAR-CLIP procedure was performed as published (Hafner et al. 2010) with the following modifications.

2.11.1 Labeling of cells with photoactivatable ribonucleosides

HEK293 T-REx Flp-In cells (Life technologies, UK) stably expressing FLAG/HA-FUS, FUS R495X, EWSR1, TAF15, TARDBP and TARDBP M337V were grown in light SILAC medium supplemented with 100 µM 4-thiouridine (4SU) (ChemGenes, USA) or 6-thioguanosine (6SG) (Sigma-Aldrich, Germany) to label long-lived transcripts. Expression of recombinant proteins was induced by addition of 200 ng/ml doxycycline (SIGMA, USA). After overnight incubation, 100 µM fresh 4SU or 6SG was added and cells were incubated for additional 2 hours to label short-lived transcripts.

2.11.2 UV crosslinking, lysis and immunoprecipitation

After aspirating the medium, cells were crosslinked on ice using a Stratalinker with customized 365 nm UV lamps (Stratagene, USA, energy settings 150 mJ/cm^2). Cells were harvested in cold PBS, pelleted by centrifugation and lysed in 3 volumes of NP40 lysis buffer. The cleared cell lysate was partially digested with 1 U/µl RNase T1 (Fermentas, Germany) for 15 min at 22°C and FLAG/HA tagged proteins were immunoprecipitated using anti-FLAG antibody at a final concentration of 0.25 µg/µl (Sigma, F1804) conjugated to Protein G Dynabeads (Invitrogen, UK). For 1 ml of cell lysate, 20 µl beads and 5 µg of antibody were used. Bound RNAs were partially digested for 6 min at 22°C with 50 U/µl RNase T1 (Fermentas, Germany). Beads were then treated with 0.5 U/µl calf intestinal phosphatase (NEB, USA) for 60 min at 37°C to dephosphorylate the RNA. Beads were washed and crosslinked RNA was labeled with 0.3 µCi/µl γ-32-P-ATP (Perkin-Elmer, NEG 502A) and 1 U/µl T4 PNK (Fermentas, Germany).

2.11.3 SDS-PAGE and electroelution

Beads were resuspended in 50 µl 2x SDS-PAGE loading buffer (20% glycerol (v/v), 160 mM Tris-HCl pH 6.8, 4% SDS (w/v), 200 mM DTT, 0.2% bromophenol blue) and RNA-protein complexes were separated by SDS-PAGE (NuPAGE Novex 4-20% BT Gel, Invitrogen, UK). The protein-RNA complex of the corresponding size was excised and electroeluted from the gel using D-Tube Dialyzer Kit MWCO 3.5kDa (Novagen, Germany) for 2h at 100V in SDS running buffer (25mM

Tris base, 192 mM glycine, 0.1% SDS). The electroeluate was digested with 2 mg/ml Proteinase K (Roche, Switzerland) for 60 min at 55°C. Immunoprecipitated RNA was recovered by phenol-chloroform extraction and ethanol precipitation.

2.11.4 RNA cloning and sequencing

Sequencing libraries were constructed using the small RNA cloning protocol (Hafner et al. 2008) with barcoded pre-adenylated 3'adapters (2.13.3.1) and sequenced on Illumina GIIA and Hiseq2000 platforms (Illumina, USA).

2.12 Chromatin Immunoprecipitation (ChIP)

2.12.1 Crosslinking and cell lysis

After siRNA transfection, cells were grown for two more days. Three million cells were used for each ChIP assay. Cells were fixed with 1% formaldehyde (Roth, Germany) for 10 min at room temperature and reaction was stopped by adding glycine at a final concentration of 125 mM for 5 min at room temperature. Cells were washed twice with ice-cold PBS + complete EDTA-free protease inhibitor cocktail (Roche, Switzerland) and harvested. Cells were resuspended in 1 ml cell membrane lysis buffer 1 (0.05 M Hepes-KOH pH 7.5, 0.14 M NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100, complete EDTA-free protease inhibitor cocktail (Roche, Switzerland)) and incubated for 15 min at 4°C with rotation. Nuclei were pelleted by centrifugation for 5 min at 4°C and 1000xg. Nuclei pellets were resuspended in 1 ml lysis buffer 2 (0.2 M NaCl, 1 mM EDTA, 0.5 mM EGTA, 10 mM Tris pH 8, complete EDTA-free protease inhibitor cocktail (Roche)), incubated for 15 min at 4°C with rotation and pelleted by centrifugation again for 5 min at 4°C and 1000xg.

2.12.2 Chromatin shearing

Pellets containing nuclei were resuspended in 300 µl S1 sonication buffer (High Cell ChIP kit, Diagenode, USA) and sonicated for 12 cycles with 30 s on and 30 s off by using a Bioruptor (UCD-300, Diagenode). The samples were centrifuged at 18,000xg for 5 min at 4°C to remove debris and detergents.

The sheared chromatin was diluted tenfold in ChIP dilution buffer (0.01% SDS, 1.1% Triton-X100, 1.2 mM EDTA, 16.7 mM Tris-HCl, pH 8.1, 167 mM NaCl, complete EDTA-free protease inhibitor cocktail (Roche, Switzerland)).

2.12.3 Immunoprecipitation

To decrease unspecific binding of chromatin to magnetic beads, chromatin was precleared by adding 25 µl of protein G magnetic beads (Invitrogen, UK) to each IP reaction and incubated at 4°C for 1 h with agitation. The supernatants were separated from magnetic beads and 10% of the precleared chromatin was kept as an input control for further analysis. 5 µg of antibody was added for immunoprecipitation at 4°C over night with rotation. The antibodies used were anti-RNA polymerase II CTD YSPTSPS antibody (ab5408, Abcam), anti-RNA polymerase II CTD YSPTSPS (phospho S2) antibody (ab5095, Abcam), anti-RNA polymerase II CTD YSPTSPS (phospho S5) antibody (ab5131, Abcam), rabbit IgG isotype control (kch-504-250, Diagenode) and mouse IgG isotype control (M5284, Sigma-Aldrich).
The next morning, after adding 25 µl of protein G magnetic beads (Invitrogen, UK), the reactions were incubated at 4°C for another 1 h with rotation.

2.12.4 Washing and elution

The beads were sequentially washed with the following washing buffers, once with low salt wash buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl, pH 8.1, 150 mM NaCl), once with high salt wash buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl, pH 8.1, 500 mM NaCl), once with LiCl wash buffer (0.25 M LiCl, 1% IGEPAL-CA630, 1% sodium deoxycholic acid , 1mM EDTA, 10 mM Tris, pH 8.1) and twice with TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0.) at 4°C.
100 µl of elution buffer (1% SDS, 0.1 M $NaHCO_3$) was added to each IP reaction with incubation at RT for 15 min. Then a second round of elution with another 100 µl of elution buffer was performed. In the end, the two supernatants were combined as the final eluate. Elution buffer was also added to input control to get a final volume of 200 µl.

2.12.5 Reverse crosslinking and DNA purification

8 µl of 5 M NaCl was added to each eluate with incubation at 65°C overnight for reverse crosslinking. The next morning, to digest RNA, 1 µl of 10 mg/ml RNase A (Roth, Germany) per reaction was added and incubated at 37°C for 30 min. Then, to digest proteins, 4 µl 0.5 M EDTA, 8 µl 1 M Tris-HCl and 1µl of 20 mg/ml Proteinase K (Roche, Switzerland) were added into each reaction and incubated at 55°C for 1 h. DNA was extracted using Phenol/Chloroform extraction by adding 400 µl of Phenol/Chloroform/Isoamyl alcohol (v/v/v=25:24:1, Roth). Then, the aqueous phase was supplemented with 400 µl of Chloroform/Isoamyl alcohol (v/v=24:1, Roth) and centrifuged at 18,000xg for 5 min at room temperature. DNA was precipitated with 1 ml of ice-cold 100% ethanol, 40 µl of 3 M sodium acetate and 1µl of Glycoblue (Ambion, USA) at -80°C for 30 min. DNA pellets were washed with 70% ethanol, air-dried and resuspended in water.

2.12.6 ChIP-qPCR and data analysis

ChIP-qPCR primers designed to amplify 50 to 150 bp DNA fragments from selected genomic regions were first evaluated by amplification efficiency (AE) evaluation with a ten-fold dilution of input DNA as qPCR template. Only primers having an AE value between 0.90-1.10 were used for subsequent qPCR experiments. Input DNA, DNA from IgG control and sample DNA were diluted three fold in water. The same volumes of DNA from above were used as qPCR templates using Power SYBR Green PCR master mix (Applied Biosystem) on the StepOne Real-Time PCR System (Applied Biosystem, USA) for 30 cycles of 15 s at 94°C, 15 s at 60°C, and 20 s at 72°C.
By visualization of amplification curves in StepOnePlus software (Applied Biosystems, USA), baseline adjustment was made if the reaction emerged before the default baseline (cycles 3 to 15). The ΔC_t value (normalized to the input samples) was calculated for each sample: ΔC_t [C_t (sample) - C_t (input)]. Then, the fold enrichment between experimental samples and input was computed by using formula $2^{(-\Delta Ct)}$.

2.13 Oligonucleotides

2.13.1 Primers

All primers were ordered from MWG Eurofins (Germany).

The following forward (_F) and reverse (_R) primers were used for PCR and cDNA cloning into pENTR4 (Invitrogen, UK), restriction sites are underlined:

FUS_F	5’-ACGCGTCGACATGGCCTCAAACGATTATACCCAAC-3’,
FUS_R	5’-ATAAGAATGCGGCCGCTCAATACGGCCTCTCCCTGCGATC-3’
FUS R495X_R	5'-ATAAGAATGCGGCCGCTTCAGAAGCCTCCACGGTCC-3'
EWSR1_F	5’-ACGCGTCGACATGGCGTCCACGGATTACAGTA-3’,
EWSR1_R	5’-ATAAGAATGCGGCCGCTTAGTAGGGCCGATCTCTGCGC-3’
TAF15_F	5’-ACGCGTCGACATGTCGGATTCTGGAAGTTACGG-3’,
TAF15_R	5’-ATAAGAATGCGGCCGCTTAGTATGGTTGCGCTGAT-3’
TARDBP_F	5’-ACGCGTCGACATGTCTGAATATATTCGGGTAACCGAAGATG
TARDBP_R	5’-TAAGAATGCGGCCGCCTACATTCCCCAGCCAGAAGACTTAG

Knockdown efficiency of proteins was analyzed by qRT-PCR using primers:

FUS qPCR_F	5’-GCCCTGGCAAGATGGATT-3’
FUS qPCR_R	5’-ACAAAAAGCTGTTCCAGAACCT-3’
EWSR1 qPCR_F	5’-AGCTACGGGCAGCAGAGTT-3’
EWSR1 qPCR_R	5’-CATGCTCCGGTTCTCTCC-3’
TAF15 qPCR_F	5’-GTCAAAACCAGCAGTCCTATCA-3’
TAF15 qPCR_R	5’-CTACTCACATCACGACGGTCA-3’
NME1-NME2 qPCR _F	5’-CCAATCCAGCAGATTCAAAG-3’
NME1-NME2 qPCR _R	5’-CATAGGCTGATTTCTTTTTCAGC-3’
PTP4A1 qPCR _F	5’-GGCCACAATCTTCAATGAGTAA-3’
PTP4A1 qPCR _R	5’-TGCTGTGCCTGGCAGTAA-3’
NUCKS1 qPCR _F	5’-ATGGTTAAGAAGTCCAAACCTG-3’,
NUCKS1 qPCR _R	5’-TTTGATGCCTTTGAAGCTGTG-3’
GAPDH qPCR _F	5’-AGCCACATCGCTCAGACAC-3’
GAPDH qPCR _R	5’-GCCCAATACGACCAAATCC-3’
STK38 qPCR_F	5’-AGACATCAAACCAGACAACCTTC-3’
STK38 qPCR_R	5’-TCCTGTGCAAAGACCAAAGTC-3’

SRSF3 qPCR_F	5’-GCCCTCGAGATGATTATCGTA-3’
SRSF3 qPCR_R	5’-CAGCGATCTCTCTCTTCTCCTATC-3’
SRSF3 qPCR_dn_F	5’-AAATGACTTGAGGGCGACAT-3’
SRSF3 qPCR_dn_R	5’-ATTGAACTGCACCCTGTGG-3’
CDKN1A qPCR_F	5’-TCACTGTCTTGTACCCTTGTGC-3’
CDKN1A qPCR_R	5’-GGCGTTTGGAGTGGTAGAAA-3’

For analysis of alternative splicing the following primers were used:

ENAH_F	5’-AGCAAGTCACCTGTTATCTCCAG-3’
ENAH_R	5’-GTCCTTCCGTCTGGACTCC-3’
THAP6_F	5’-GAGATGTGTTGTGTTCGAGGC-3’
THAP6_R	5’-CCTTTGTATCCTCTAGCTCGC-3’
CSNK1D_F	5’-CGTCAACATCTCCTCGTCC-3’
CSNK1D_R	5’-GCACGACAGACTGAAGACC-3’
PDE8A_F	5’-GGCTTGTAACTCAGTATTCACTGC-3’
PDE8A_R	5’-TCCAATGACAGGTATTATCTTCACA-3’

The following primers were used for ChIP-qPCR:

SRSF3 TSS_F	5’-AGGCGGTGGTCCGCCATTTC-3’
SRSF3 TSS_R	5’-CCGCTTTCCTCCGGCCCAAC-3’
GAPDH TSS_F	5’-TACTAGCGGTTTTACGGGCG-3’
GAPDH TSS_R	5’-TCGAACAGGAGGAGCAGAGAGCGA-3’
SRSF3 dn1_F	5’-CACAGAGGGATGACCGTGT-3’
SRSF3 dn1_R	5’-TCTGTCCCTGCTTGCAGAC-3’
SRSF3 dn2_F	5’-ACCGTGTGAGGAGGCAGTAG-3’
SRSF3 dn2_R	5’-ACCGTGTGAGGAGGCAGTAG-3’
SRSF3 dn4_F	5’-GGCTCCGCTTTCTCAGAGTT-3’
SRSF3 dn4_R	5’-AGCCTGTGGTGATGGTGATG-3’
SRSF3 3'UTR_F	5’-GAAACACAGGCCATCAGGGA-3’
SRSF3 3'UTR_R	5’-ACCAACTAGGCAACCTCTGC-3’
GAPDH 3’UTR_F	5’-CCCCCACCACACTGAATCTC-3’
GAPDH 3’UTR_R	5’-TGGTTGAGCACAGGGTACTT-3’
Untr control1_F	5’-AAGTTATCATCCTGGTGAGTTGC-3’
Untr control1_R	5’-AGGTAATTAATCTGCTACTCTGGGA-3’

The following primers were used for PAR-CLIP:

3’ PCR primer	5’-CAAGCAGAAGACGGCATACGA-3‘
5’ PCR primer	5‘-AATGATACGGCGACCACCGACAGGTTCAGAGTTCTAC AGTCCGA-3‘

2.13.2 siRNAs

The following siRNA duplexes (sense/antisense) were used for knockdown experiments. TAF15 siRNAs were designed as described by (Jobert et al. 2009). All siRNA duplexes were purchased from Sigma-Aldrich (Germany):

FUS duplex 1, GAUCAAUCCUCCAUGAGUAdTdT, UACUCAUGGAGGAUUGAUCdTdT,
FUS duplex 2, CAGAGUUACAGUGGUUAUAdTdT, UAUAACCACUGUAACUCUGdTdT,
EWSR1 duplex 1, GACUCUGACAACAGUGCAAdTdT, UUGCACUGUUGUCAGAGUCdTdT,
EWSR1 duplex 2, GCCAAGCUCCAAGUCAAUAdTdT, UAUUGACUUGGAGCUUGGCdTdT,
TAF15 duplex 1, UGAUCAGCGCAACCGACCAdTdT, UGGUCGGUUGCGCUGAUCAdTdT,
TAF15 duplex 2, GGACAGAACUACAGCGGUUdTdT, AACCGCUGUAGUUCUGUCCdTdT

2.13.3 Adapters

2.13.3.1 PAR-CLIP adapters

5’adapter

oR5-NN	5’-GUUCAGAGUUCUACAGUCCGACGAUCNN-3‘

3’adapters

NN-NBC1	5’-NNAAAATCGTATGCCGTCTTCTGCTTG-3’
NN-NBC2	5’-NNCCCATCGTATGCCGTCTTCTGCTTG-3’
NN-NBC3	5’-NNGGGATCGTATGCCGTCTTCTGCTTG-3’
NN-NBC4	5’-NNTTTATCGTATGCCGTCTTCTGCTTG-3’
NN-NBC5	5’-NNCACGTCGTATGCCGTCTTCTGCTTG-3’
NN-NBC6	5’-NNCCATTCGTATGCCGTCTTCTGCTTG-3’
NN-NBC7	5’-NNCGTATCGTATGCCGTCTTCTGCTTG-3’
NN-NBC8	5’-NNCTGCTCGTATGCCGTCTTCTGCTTG-3’

2.13.3.2 Small RNA cloning adapters

5'adapter

oR5-NN 5'-GUUCAGAGUUCUACAGUCCGACGAUCNN-3'

3'adapters

NBC1 5'-TCTaaaaGTATGCCGTCTTCTGCTTGT-3'

NBC2 5'-TCTcccaGTATGCCGTCTTCTGCTTGT-3'

NBC3 5'-TCTgggaGTATGCCGTCTTCTGCTTGT-3'

NBC4 5'-TCTtttaGTATGCCGTCTTCTGCTTGT-3'

NBC5 5'-TCTcacgGTATGCCGTCTTCTGCTTGT-3'

NBC6 5'-TCTccatGTATGCCGTCTTCTGCTTGT-3'

NBC7 5'-TCTcgtaGTATGCCGTCTTCTGCTTGT-3'

NBC8 5'-TCTctgcGTATGCCGTCTTCTGCTTGT-3'

2.14 Computational methods

Data analysis as described below was conducted by Marvin Jens, PhD student in the group of Nikolaus Rajewsky at the Berlin Institute for Medical Systems Biology at the Max-Delbrück Center for Molecular Medicine (Robert-Rössle Str. 10, D-13125 Berlin, Germany)

2.14.1 PAR-CLIP computational pipeline

All sequencing data sets were run with a pipeline that performed all steps of the analysis from raw reads to cluster sets and target genes, in a largely automated and unbiased way. The emphasis was on stringent filtering and controlling the false-positive rate in the identification of binding sites. Reads were collapsed into distinct sequences (counting each sequence only once) and aligned to the reference genome assembly hg18 allowing for up to one mismatch, insertion or deletion. Only uniquely mapping reads were retained. Clusters of aligned CLIP-seq reads were identified that continuously covered regions of genomic sequence.

A number of additional quality scores to each cluster were also assigned, based on read coverage, the relative frequency of characteristic conversions and uniqueness of alignments:

- number of (unique, distinct) read alignments
- number of characteristic mismatches (T to C, G to A,...)
- length of the cluster
- entropy score over read start/end positions
- entropy score over read sequence variability
- maximum *uniqueness* of all alignments that support a cluster

Uniqueness refers to the margin between the reported, best alignment of a read and the second best alignment considered by the read mapper.
As the aggregate amount of sense and antisense sequence is identical (ambiguous cases are put aside), one can regard the reverse complement of all transcripts, as an approximately fair decoy database. In the absence of any real biological signal in the PAR-CLIP data an equal number of clusters to hit sense and antisense may be expected. Such a decoy database, therefore introduces a simple estimator of the false discovery rate (FDR) in the set of all PAR-CLIP read clusters:

FDR = (#antisense + 1)/(#sense + #antisense + 2)

where #sense and #antisense refer to the number of sense and antisense clusters, respectively.
With the FDR estimator at hand, it is possible to assess the effect of filtering the cluster set by setting thresholds on their quality scores. If the antisense clusters indeed represent mapping artifacts, the corresponding quality score distribution should differ from the sense clusters, which supposedly contain the true-positives. This would allow finding cutoffs that deplete false-positives more strongly than true-positives and improve the FDR. It is important to bear in mind, that mapping artifacts may also align sense to known transcripts. Utilizing the antisense clusters to select cutoffs will arguably serve to also deplete the false-positives among the sense aligning clusters. On the other hand, the filtering should discard as little real data as possible. To find the best compromise, the pipeline code iterates over each of the aforementioned cluster quality scores and estimates the FDR at each quality score that actually appears in the data, effectively probing the whole range of possible cutoffs.

If a score cutoff serves to reduce the FDR below a desired limit (FDR < 5%) it is recorded, together with the number of sense clusters that surpass the cutoff and would be retained. Out of all score/cutoff combinations that satisfy the FDR limit, the one preserving the largest number of sense aligning clusters is chosen. After the cutoff is applied, remaining decoy clusters are discarded and a cluster set is reported that can be expected to satisfy the FDR constraint.
In order to screen for evidence of potential interactions between FET proteins and splice sites, we performed a reciprocal analysis. Taking a particular set of processing sites (e.g.3' splice sites) we scanned for the presence or absence of PAR-CLIP coverage in the neighbourhood of all sites. The windows of zeros and ones (indicating absence or presence of at least one aligned read) were accumulated and averaged to yield the average probability for finding at least one PARCLIP read at a given distance.

2.14.2 Sequence motif analysis

6-mer occurrences were counted in 41nt windows around preferred crosslink sites identified in the 4SU and 6SG PAR-CLIP experiments. 6-mers with less than 10 occurrences were discarded from the analysis. The frequency of the remaining 6-mers was compared to all reference sequences (RefSeq) 3'UTR sequences or introns as a representative background set (Lebedeva et al. 2011).

2.14.3 RNA secondary structure analysis

To test whether FET binding sites showed a preferred secondary structure the library routines from the Vienna RNA package 1.8.2 (Hofacker 2004) were used to compute base pairing probabilities within 201nt sequences centered on the preferred crosslink positions of binding sites. The resulting profiles were accumulated and averaged over all sites. Randomly chosen positions served as a control.

2.14.4 Gene ontology term analysis

Target genes were subjected to GO term analysis using the web-based DAVID functional annotation tool (Huang et al., 2009 1 and 2).

2.14.5 RNA-Seq quantification

Polyadenylated RNA from mock-transfected cells and siRNA treated cells were sequenced on Illumina Genome Analyzer GAII or Illumina HiSeq (Illumina, USA) using 2x76 bp paired-end kits.

All obtained paired-end reads were mapped to the hg18 genome sequence (Pruitt et al. 2005) with tophat (Trapnell et al. 2009). The quantification utility CUFFDIFF from the CUFFLINKS RNA-Seq toolchain (Trapnell et al. 2010) was used to estimate gene and isoform FPKM-levels and confidence intervals using the aligned reads and the RefSeq gene models. Log2 fold changes were computed from the inferred gene-level FPKM values from siRNA and mock-transfected samples. We computed log fold changes only for genes that had >= 5 FPKM in either of the compared samples and added a pseudo-count of 10 FPKM to reduce noise from low expression and avoid divisions by zero.

2.14.6 3'UTR extension analysis

A small RNA library from RNase-I digested poly(A)+ mRNA (provided by M. Munschauer), was utilized to unbiasedly annotate transcribed regions in the genome of HEK293T-Rex cells with strand information. Regions that contained more than 10 aligned reads within a kilobase were merged into "islands" and annotated against the RefSeq annotation [RefSeq consortium, BLAT, UCSC genome browser]. In this manner, 3' extensions of known transcription units were identified by strand and proximity to or overlap with known sites of cleavage and polyadenylation. Such 3'UTR extension regions were then scored for the number or polyA mRNA-Seq reads aligning to them in unperturbed (mock) and FET-protein loss of function conditions (RNAi knockdown) to identify regions of putative FET-protein dependent transcription.

3. Results

3.1 Generation of cell lines for stable, inducible expression of epitope-tagged FET proteins

To identify RNA and DNA elements interacting with FUS, EWSR1 and TAF15, HEK293 cell lines stably inducible expressing N-terminal FLAG/HA-tagged versions of the proteins were generated. The FLAG/HA-tag of the proteins was then used to efficiently immunoprecipitate the recombinant proteins for further experiments (PAR-CLIP and ChIP). Likewise, a cell line for the FUS variant R495X found in patients with ALS was generated for comparison. This mutation lacks a putative nuclear localization signal at the C-terminus of FUS.

Cell lines were generated by cloning and recombination the respective coding sequence of FUS, EWSR1, TAF15 and the FUS mutant R495X using the FLP-In T-Rex System (Invitrogen). Expression of the recombinant proteins is controlled by a Tet operator allowing for specific induction of protein expression after addition of doxycycline. The epitope-tagged proteins were induced to expression levels lower than the respective endogenous proteins as observed by Western blot analysis (Fig.5) using protein-specific antibodies to detect endogenous (lower band) as well as FLAG/HA-tagged FET proteins (upper band).

This indicates that the expression level of recombinant FET proteins is comparable with the level of endogenous FET proteins.

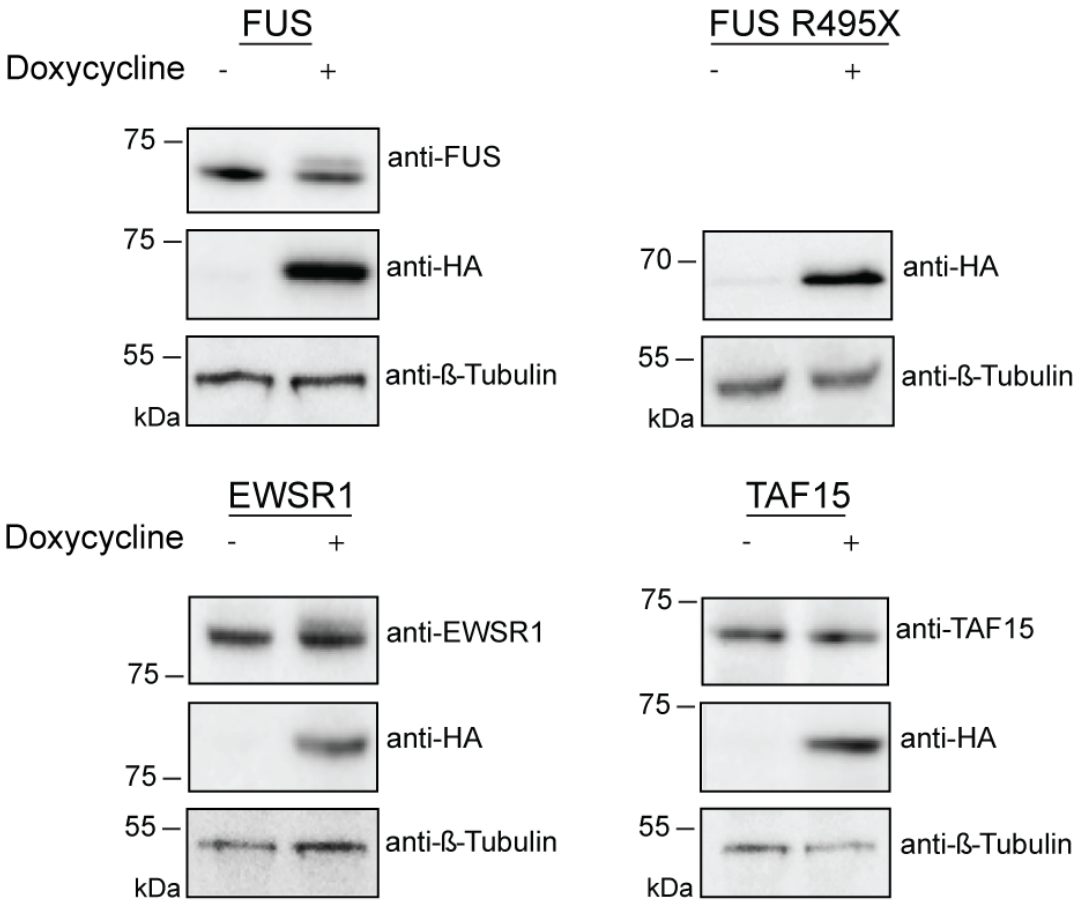


Fig. 5: Generated cell lines for stable, inducible expression of FLAG/HA-tagged FET proteins.
Western blot analysis of cell lysates before and after induction of recombinant protein expression with doxycycline. Samples were analyzed by Western blotting with antibodies against FUS, EWSR1, TAF15 (upper panels) and HA-epitope (middle panels). Due to the C-terminal mutation FUS R495X cannot be detected by anti-FUS. ß-Tubulin was detected as a loading control (lower panels).

3.2 Identification of FET protein RNA targets by PAR-CLIP

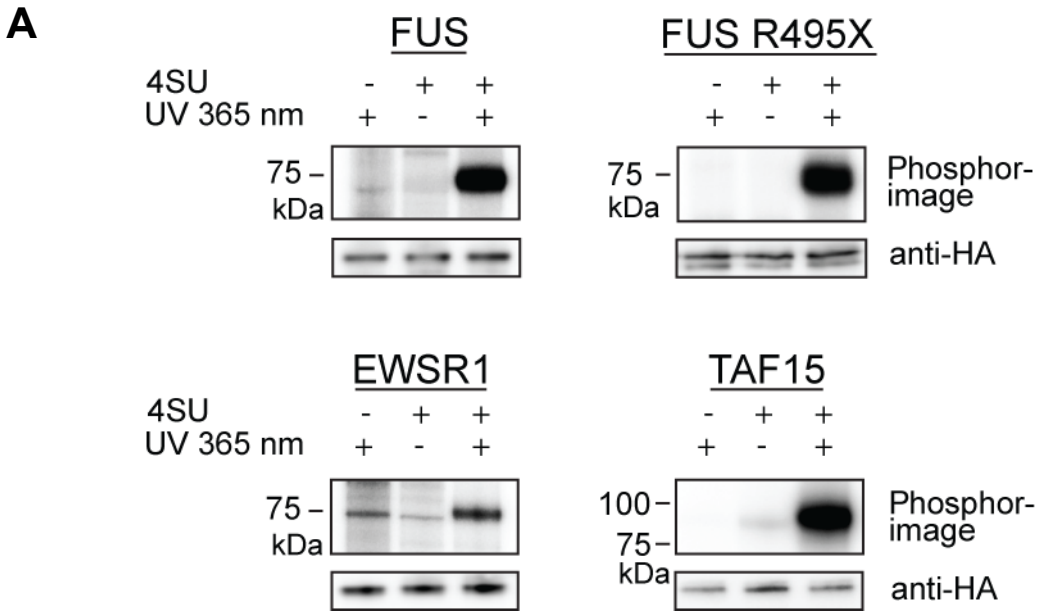
To identify RNA elements interacting with FUS, FUS R495X, EWSR1 and TAF15 the recently developed PAR-CLIP method was applied followed by next-generation sequencing (Hafner et al. 2010) (Tab.2). During the PAR-CLIP experiments cellular RNA is labeled with either 4-thiouridine (4SU) or 6-thioguanosine (6SG) and crosslinked to bound proteins (Fig.6A). Efficient crosslinking leads to specific nucleotide transition events during the reverse transcription step in library preparation from each experiment: crosslinked 4SU and 6SG residues are converted into C and A, respectively (Fig.6B), providing a diagnostic mark at nucleotide resolution of the RBP binding site on target RNAs (Hafner et al. 2010).

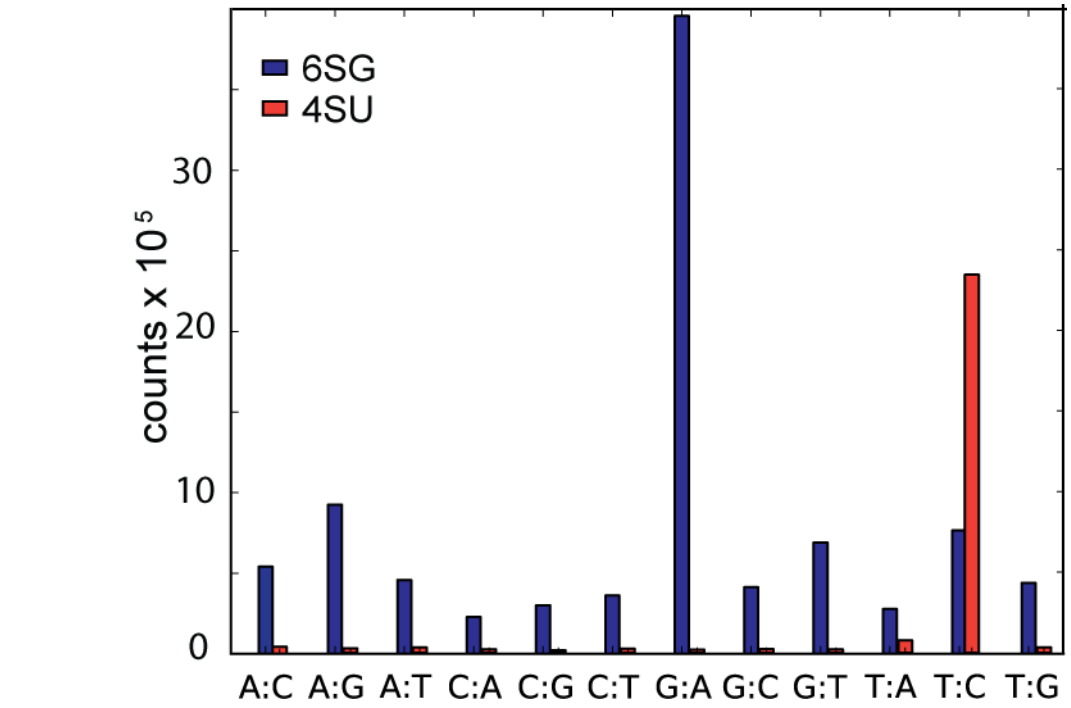
Tab. 2: Overview of PAR-CLIP experiments and samples.

	4-SU PAR-CLIP libraries	6-SG PAR-CLIP libraries
FUS	4x	1x
FUS R495X	1x	1x
EWSR1	2x	
TAF15	2x	1x
TARDBP	2x	1x
TARDBP M337V	1x	

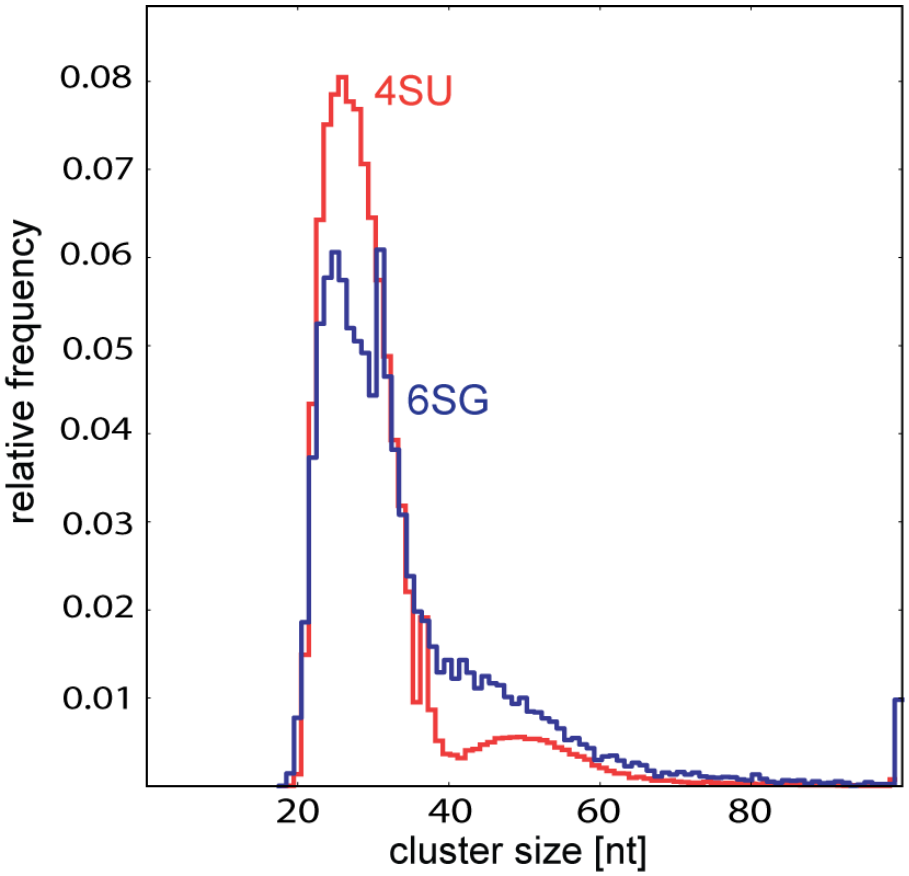
All PAR-CLIP sequencing data were analyzed with a recently described computational pipeline (Lebedeva et al. 2011) to determine the binding sites at an estimated 5% FDR from filtered clusters of aligned reads. To define a PAR-CLIP consensus, clusters that were not supported by reads from at least 3 out of 5 FUS libraries, 2 out of 2 EWSR1 libraries, 2 out of 3 TAF15 libraries, 2 out of 3 TARDBP libraries and 2 out of 2 FUSR495X libraries (biological replicates) were discarded. The number of T to C or G to A mismatches (4SU, 6SG characteristic conversions) served as a crosslink score (Hafner et al. 2010). As PAR-CLIP cDNA reads are typically short and their sequence is mutated by 4SU-induced crosslinking, they cannot be expected to always align correctly to the reference sequence. A certain fraction will produce false alignments, leading to false-positive binding sites. However, as the bound RNA fragments derive from biological, naturally occurring RNA and the sequencing strategy preserves strand information, the true-positives can be expected to align predominantly sense to known transcripts. Clusters aligning antisense to known transcripts, on the other hand, can be regarded as false-positives. In a few cases they may be true-positives, derived from un-annotated antisense transcripts, but these will be rare and typically much less abundant. Stranded sequencing data was used to comprehensively annotate regions of transcription, adding significant amounts of apparently true antisense transcripts to the RefSeq catalog. Consequently, treating remaining antisense aligning clusters as false-positives, is a conservative assumption because it can only over estimate the number of false-positives produced by alignment artifacts.

The numbers of obtained sequence clusters ranged from 55.000 to 232.000 for the four proteins targeting 6.591 to 11.470 genes which encode up to 78% of all proteins expressed in HEK293 cells. Figure 6D shows an example of aligned reads and transition events for the FUS target *CDK1*. A mean length of 25 nt for filtered FUS, EWSR1, and TAF15 sequence clusters demonstrates the high-resolution of PAR-CLIP data (Fig.6C).


A
FUS
FUS R495X
4SU
UV 365 nm
75 –
kDa
Phosphor-
image
anti-HA
EWSR1
TAF15
100 –
75 –


B
6SG
4SU
counts x 10^5
30
20
10
0
A:C A:G A:T C:A C:G C:T G:A G:C G:T T:A T:C T:G

C


4SU
6SG
relative frequency
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
20
40
60
80
cluster size [nt]

D

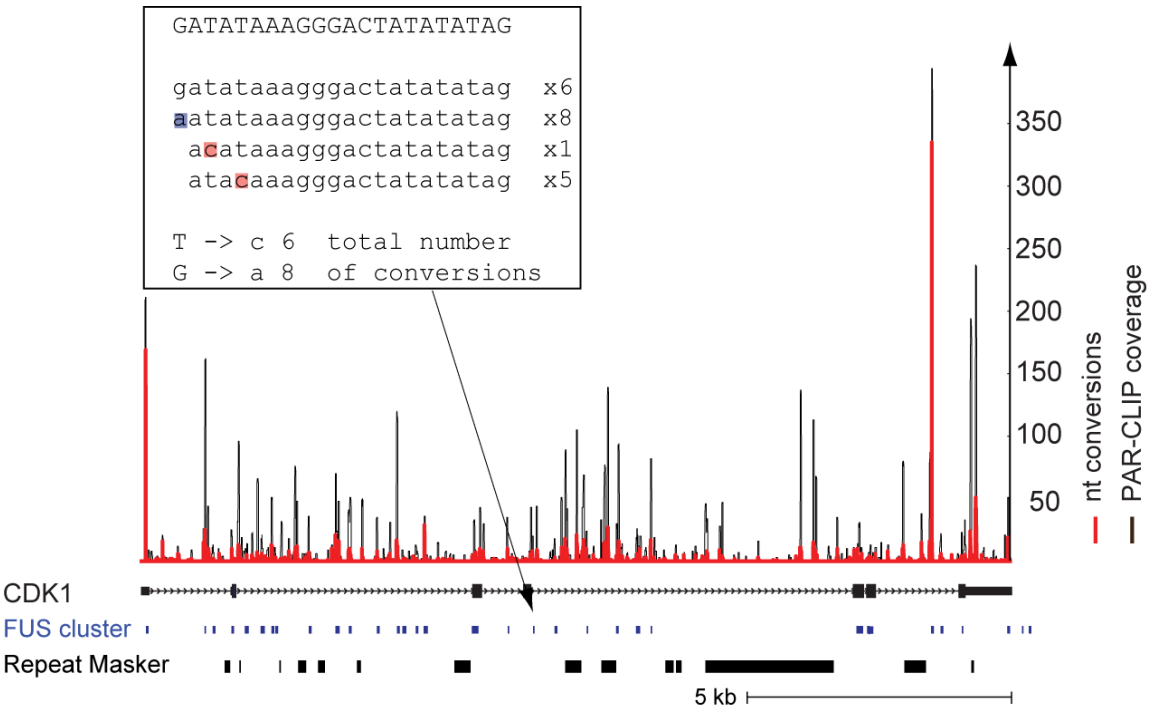

GATATAAAGGGACTATATATAG
gatataaagggactatatatag x6
aatataaagggactatatatag x8
acataaagggactatatatag x1
ataccaaagggactatatatag x5
T -> c 6 total number
G -> a 8 of conversions
350
300
250
200
150
100
50
nt conversions
PAR-CLIP coverage
CDK1
FUS cluster
Repeat Masker
5 kb

Fig. 6: PAR-CLIP of FET proteins.

(**A**) Efficient UV 365 nm crosslinking of 4SU-containing RNA to FET proteins. Phosphoimages of SDS-PAGE gels resolving radiolabeled RNA- FET protein-immunoprecipitates are shown in the upper panel. Expression of FLAG/HA-tagged FET proteins was induced by doxycycline and cells were cultured in the absence (-) or presence (+) of 4SU. Equal amounts of immunoprecipitated protein was confirmed by Western blot analysis using an anti-HA antibody (lower panel). (**B**) Crosslinking of 4SU or 6SG-labeled RNA resulted in specific mismatches during RT-PCR which are characterized by T to C and G to A conversions, respectively. The frequency of nucleotide mismatches in PAR-CLIP reads is shown for 4SU (red) and 6SG (blue). (**C**) Example of length distribution of PAR-CLIP clusters for a 4SU (red) and 6SG library (blue). (**D**) Representative example of PAR-CLIP data. The read coverage (black) and nucleotide conversions (red) are given together with FUS binding sites (blue boxes) and repetitive elements (black boxes) for the *CDK1* transcript. Sequences which cannot be mapped uniquely due to repetitive elements are discarded. Insert: Example of a FUS cluster with T to C conversions highlighted in red and G to A conversions highlighted in blue.

RBP binding sites are defined in the PAR-CLIP computational pipeline as clusters of overlapping or directly adjacent sequencing reads which contain characteristic nucleotide conversions. Uniquely aligning reads were grouped into read clusters. Clusters mapping antisense to known transcripts were used to estimate the FDR. Clusters overlapping repetitive elements were discarded.

The majority of clusters identified for the three FET proteins mapped to transcripts derived from protein-coding genes. About 90% of the FUS and EWSR1 mRNA-mapping clusters aligned to intronic sequences, consistent with the nuclear localization of these two proteins (Andersson et al. 2008), whereas TAF15 sequence clusters were nearly equally distributed in introns and 3'UTRs (Fig.7A). As expected, for the FUS variant R495X, which localizes predominantly to the cytoplasm due to the loss of the C-terminal nuclear localization signal (Fig.7B), a change in the distribution of binding sites was observed compared to the wild-type protein resulting from a reduction in intronic binding sites likely due to its loss of the NLS.

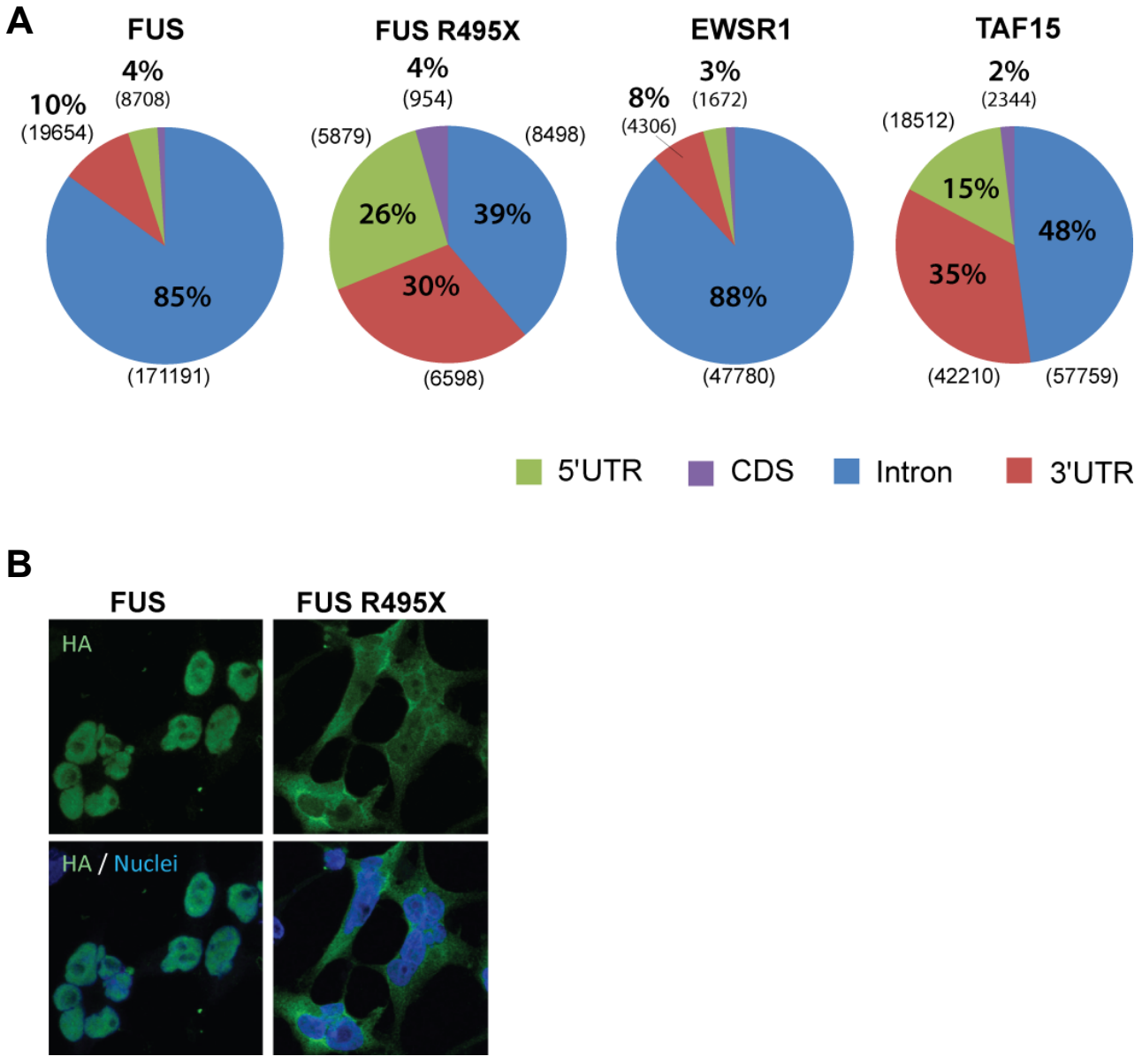


Fig. 7: Comparison of FET and FUS R495X binding sites.

(**A**) Distribution of FET and FUS R495X binding sites along transcripts. For EWSR1 and FUS most of the binding sites are located in introns. For TAF15 a larger proportion of binding sites are also located in 3'UTRs. Due to the altered cytoplasmic localization of FUS R495X, the number of intronic binding sites is reduced compared to FUS. (**B**) Differences in localization of wildtype and FUS R495X mutant. FLAG/HA-tagged FUS and FUS R495X expressing HEK293 cells were stained with an anti-HA antibody (green), the nuclear counter stain TO-PRO-3 (blue) and analyzed by microscopy. Images were provided by Dr. Dorothee Dormann, DZNE, Ludwig-Maximilians-Universität München.

3.3 Identification of TARDBP RNA targets and binding sites

In 2006, TARDBP/TDP-43 was identified as a major component of ubiquitinated inclusions in the central nervous system of patients with ALS and FTLD (Neumann et al. 2006). Shortly after, mutations in TARDBP were discovered as causative of ALS and FTLD cases (Neumann et al. 2006; Kabashi et al. 2008; Sreedharan et al. 2008). Mutations in a second RBP, FUS, were identified in familial ALS patients as well as in rare FTLD cases (Neumann et al. 2009; Vance et al. 2009). TARDBP is structurally unrelated to FET proteins, but likewise, ubiquitinated cytoplasmic inclusions containing TARDBP are observed in disease-affected tissues (Neumann et al. 2006). These genetic findings and the common pathology indicate that both FUS and TARDBP proteins abnormally aggregate in ALS and FTLD, and suggest similar RNA targets and molecular mechanisms aberrantly regulated at the post-transcriptional level as potential pathogenic clues.

For comparison, cell lines of TARDBP and the ALS related TARDBP mutant M337V were generated and PAR-CLIP was performed (Fig.8A). The genomic annotation of the TARDBP sequence clusters is consistent with previous reports of TARDBP binding patterns in embryonic stem cells, SH-SY5Y neuroblastoma cells and human brain (Tollervey et al. 2011; Xiao et al. 2011). In comparison with PAR-CLIP data of the FET proteins, close to 70% of TARDBP sequence clusters were found in introns of protein-coding transcripts (Fig.8B), in agreement with HITS-CLIP and iCLIP data (Polymenidou et al. 2011; Tollervey et al. 2011). Even more intronic binding sites were found for the TARDBP mutant M337V, which in contrast to the FUS R495X mutant shows no altered localization in the cell compared to the wild type protein (Fig.8C).

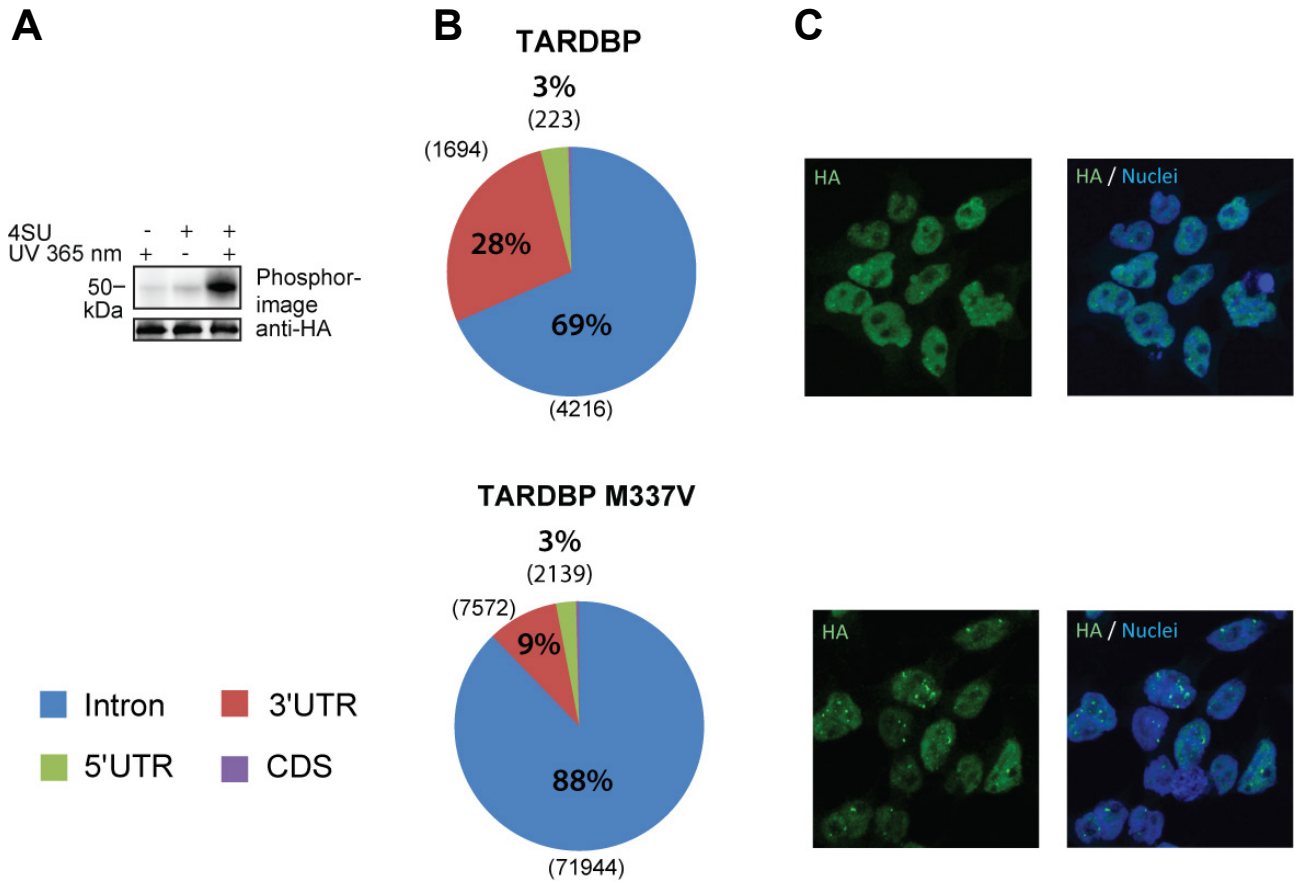


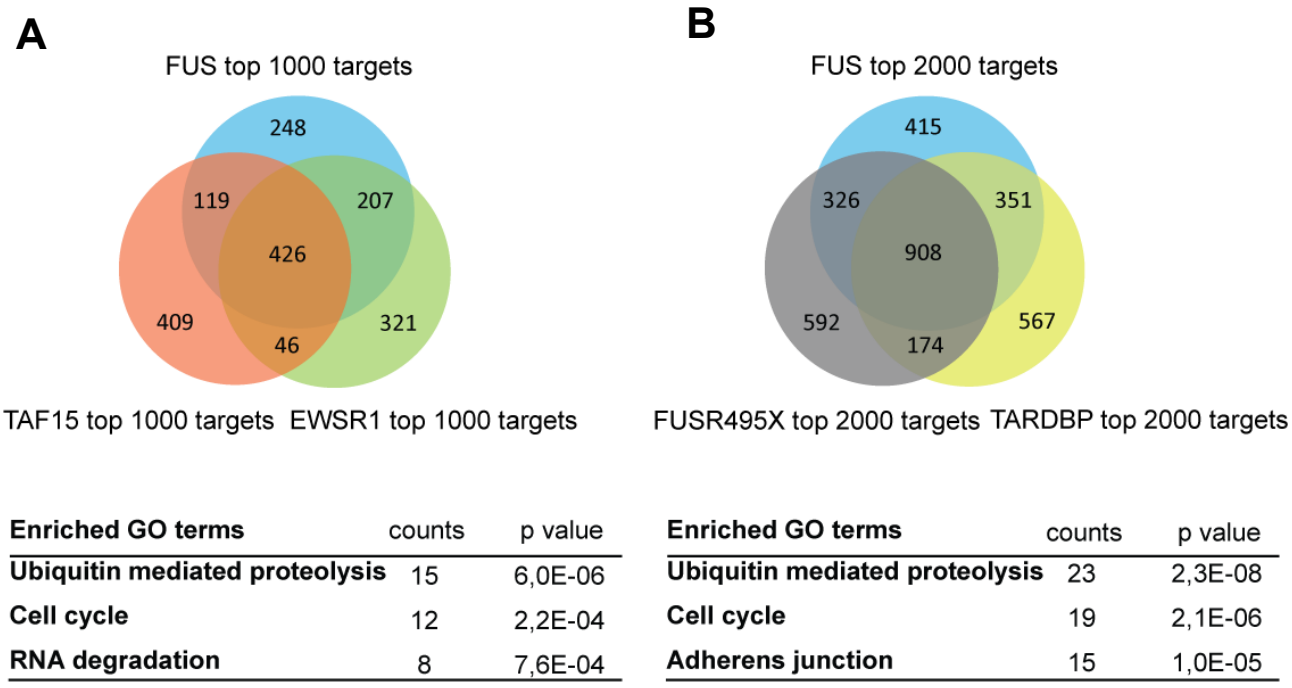
Fig. 8: Comparison of TARDBP and TARDBP M337V binding sites.

(**A**) Phosphoimage of SDS-gel resolving radiolabeled RNA-TARDBP immunoprecipitates are shown in the upper panel. Expression of FLAG/HA-tagged TARDBP was induced by doxycycline and cells were cultured in the absence (-) or presence (+) of 4SU. Equal immunoprecipitation was confirmed by Western blot analysis using anti-HA (lower panel). (**B**) Distribution of TARDBP and TARDBP mutant M337V binding sites along transcripts. (**C**) Nuclear localization of wildtype and TARDBP M337V mutant. FLAG/HA-tagged TARDBP and TARDBP M337V expressing HEK293 cells were stained with an anti-HA antibody (green), a nuclear counter stain (blue) and analyzed by microscopy. Images were provided by Dr. Dorothee Dormann, DZNE, Ludwig-Maximilians-Universität München.

3.4 Overlap of FET and TARDBP mRNA targets

In order to obtain insight whether FUS, EWSR1 and TAF15 regulate different sets of transcripts, the top 1000 target mRNAs ranked by T-to-C frequency were examined (Fig.9A). This comparison revealed that FUS and EWSR1 overlap in 643 targets and all three family members share 426 mRNA targets. Gene ontology (GO) term analysis (Huang da et al. 2009b; Huang da et al. 2009a) of mRNAs bound by all three FET proteins revealed an enrichment of the GO terms “ubiquitin mediated proteolysis”, “cell cycle” and “RNA degradation”. Furthermore, mRNA targets of the ALS related protein TARDBP also overlap substantially with FUS and FUS R495X targets.

Intersection of the top 2000 FUS, FUS R495X and TARDBP targets revealed 908 overlapping targets (Fig.9B). These target genes were subjected to GO term enrichment analysis. The most enriched category suggests also an involvement of FET proteins and TARDBP in the regulation of the ubiquitin-mediated proteolysis pathway.



Enriched GO terms	counts	p value
Ubiquitin mediated proteolysis	15	6,0E-06
Cell cycle	12	2,2E-04
RNA degradation	8	7,6E-04

Enriched GO terms	counts	p value
Ubiquitin mediated proteolysis	23	2,3E-08
Cell cycle	19	2,1E-06
Adherens junction	15	1,0E-05

Fig. 9: Comparison of FET and TARDBP mRNA targets.

(**A**) Overlap of top 1000 mRNA targets of FUS, EWSR1 and TAF15, together with enriched GO terms among FET overlapping targets. (**B**) Overlap of top 2000 mRNA targets of FUS, FUS R495X and TARDBP with GO term analysis of overlapping targets. GO term analysis was conducted with the DAVID online tool (Huang da et al. 2009b).

3.5 FUS and TAF15 recognize different motifs than EWSR1

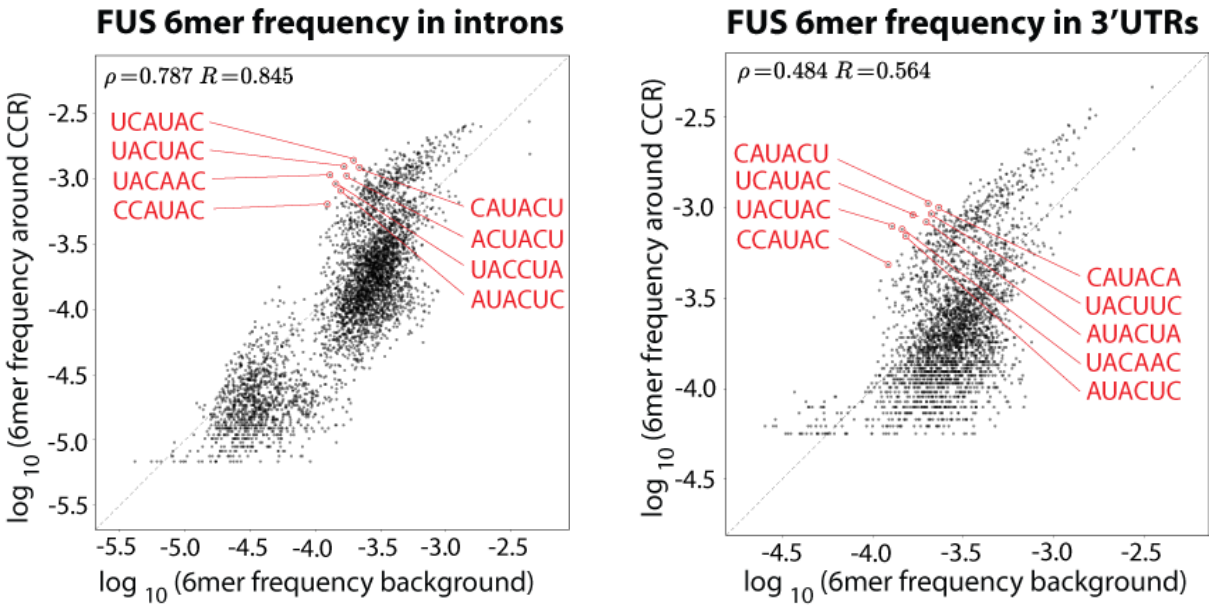
Next, 4SU and 6SG PAR-CLIP data were used to identify RNA sequences that were preferentially recognized by FET proteins. Since only uridine (U) containing binding sites can be captured by using 4SU in PAR-CLIP experiments, the 6SG PAR-CLIP data were used to capture also sequences devoid of Us. Hence, the 6mer occurrence in 41 nt window centered on the anchor was counted. An anchor corresponds to the position in a cluster with the highest frequency of nucleotide transitions and therefore represents the preferred binding site (crosslink centered region, CCR).

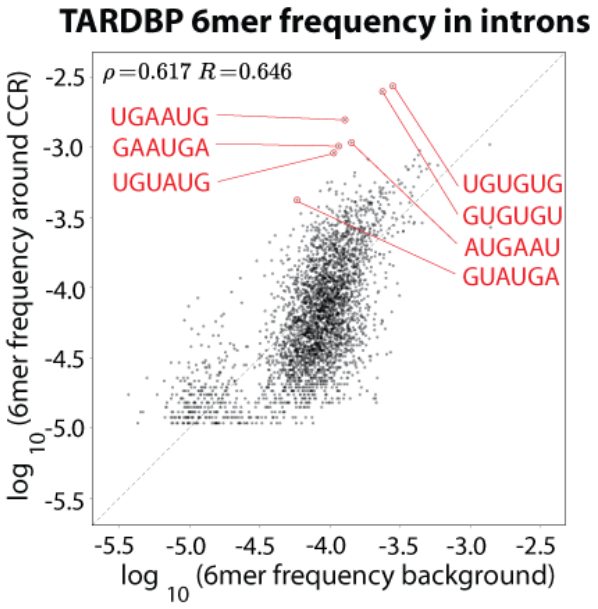
Among the most abundant words preferentially bound by the FET proteins, UAC containing sequences are clearly enriched for FUS and TAF15 (Fig.10A). Interestingly, for EWSR1 GC-rich, but no UAC-rich sequences, were enriched among the top motifs.

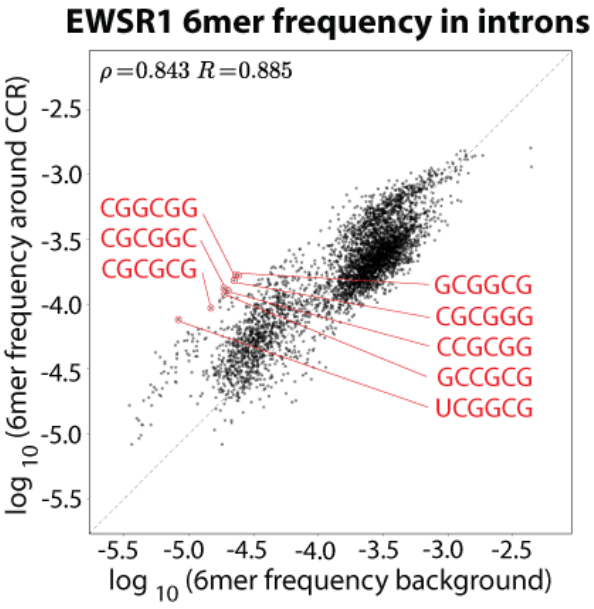
We analyzed our TARDBP PAR-CLIP data the same way and counted the 6mer occurrence around the anchor. The most significantly enriched hexamer was UGUGUG (Fig.10A) as recently discovered by iCLIP (Tollervey et al. 2011).

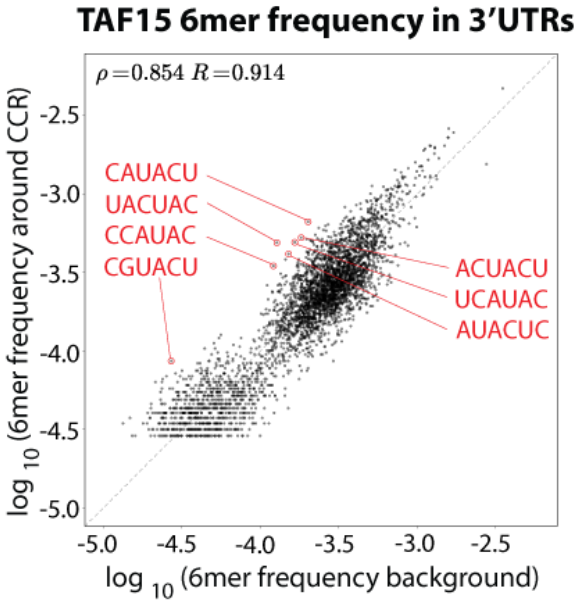
Furthermore, the PAR-CLIP data were used to investigate whether FET binding sites have any specific predicted secondary structure. RNA secondary structure can be a feature contributing to the specificity of RBPs (Buratti and Baralle 2004). The FET proteins have been proposed to bind to AU-rich stem-loop structures (Hoell et al. 2011). Therefore, the binding sites of FET proteins in introns were computationally folded and the resulting base pairing probabilities were averaged. We found a substantially reduced base-pairing probability in the direct vicinity of FUS, EWSR1 and TAF15 anchors with no indication of a hairpin structure (Fig.10B).

A




TARDBP 6mer frequency in introns
ρ=0.617 R=0.646
UGAAUG
GAAUGA
UGUAUG
UGUGUG
GUGUGU
AUGAAU
GUAUGA
log 10 (6mer frequency around CCR)
log 10 (6mer frequency background)
-2.5
-3.0
-3.5
-4.0
-4.5
-5.0
-5.5


EWSR1 6mer frequency in introns
ρ=0.843 R=0.885
CGGCGG
CGCGGC
CGCGCG
GCGGCG
CGCGGG
CCGCGG
GCCGCG
UCGGCG
log 10 (6mer frequency around CCR)
log 10 (6mer frequency background)
-2.5
-3.0
-3.5
-4.0
-4.5
-5.0
-5.5


TAF15 6mer frequency in 3'UTRs
ρ=0.854 R=0.914
CAUACU
UACUAC
CCAUAC
CGUACU
ACUACU
UCAUAC
AUACUC
log 10 (6mer frequency around CCR)
log 10 (6mer frequency background)
-2.5
-3.0
-3.5
-4.0
-4.5
-5.0

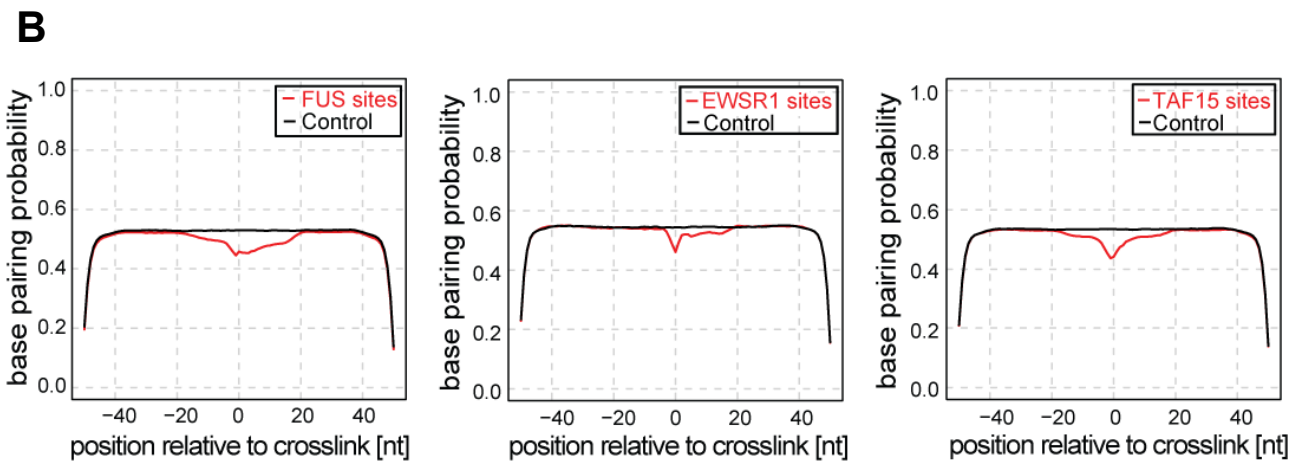


Fig. 10: FET binding motif preferences.

(**A**) Correlation between 6mer frequencies in FET PAR-CLIP experiments versus matched control regions. Pearson and Spearman rank correlation coefficients are indicated. (**B**) FET proteins bind preferentially single stranded regions. 41 nt sequence contexts around FET binding sites were computationally folded and the average base pairing probability was determined.

3.6 Knockdown of FET proteins reduces target mRNA abundance

To test whether the large number of FET interactions uncovered by PAR-CLIP were functional and to study the influence of the FET family on gene expression, FUS, EWSR1 and TAF15 were depleted in HEK293 cells by siRNA mediated knockdown using two different siRNAs for each protein. Since FUS, EWSR1 and TAF15 may be redundant, all three proteins were depleted at the same time (triple knockdown) as well.

Knockdown efficiency of different siRNAs was validated by qRT-PCR and Western blot (Fig.11). FUS, EWSR1 and TAF15 are abundant proteins and therefore a high efficiency of the knockdown was required to achieve a reproducible and effective response. The knockdown efficiency quantified by qRT-PCR ranged between 3 to 5 fold for the three FET proteins.

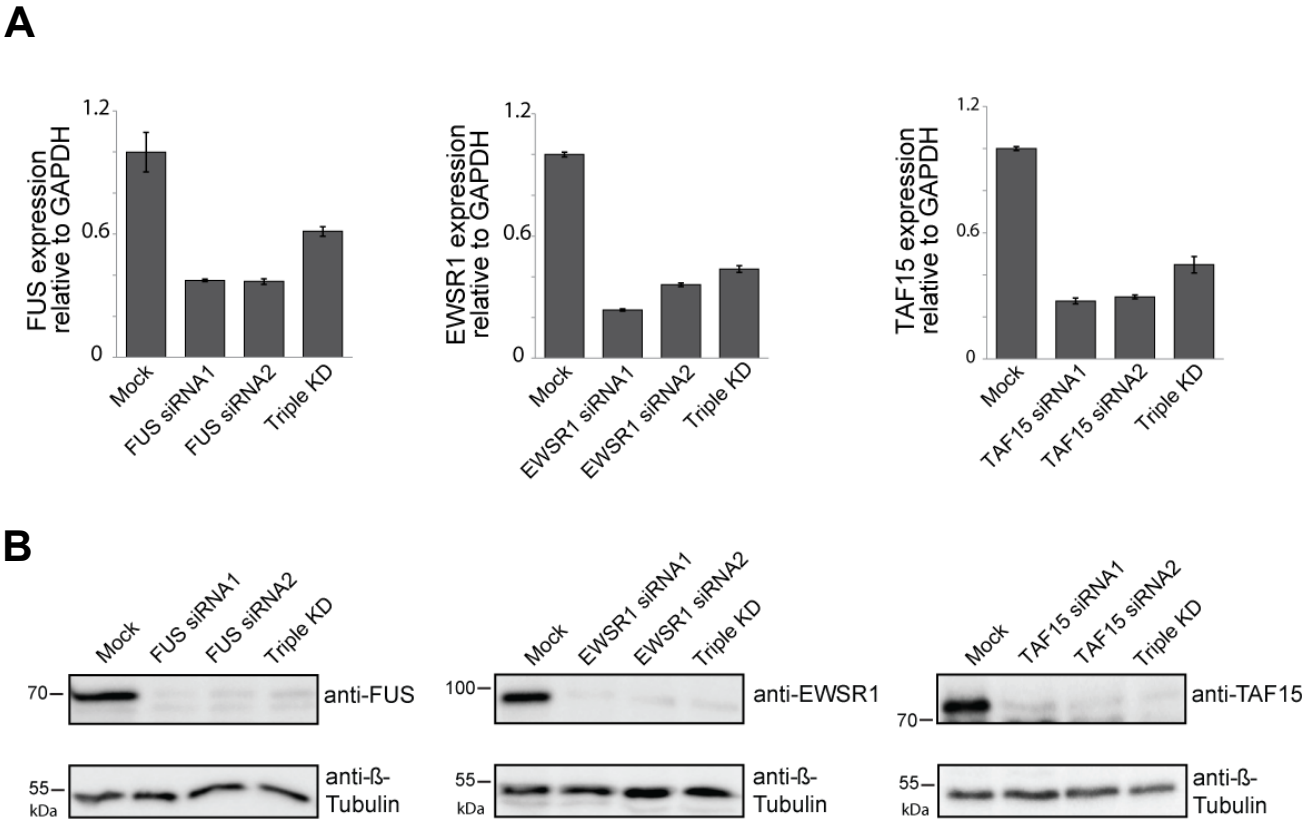


Fig. 11: Validation of FET knockdown.

(**A**) Quantitative RT-PCR on total RNA from cells after FET knockdown using GAPDH for normalization. Error bars represent the standard deviation for three technical replicates. (**B**) Cell lysates of FUS, EWSR1, TAF15 siRNA treated and untreated cells (mock) were immunoblotted and probed with anti-FUS, anti-EWSR1, anti-TAF15 (upper panels) and anti-β-Tubulin (lower panels).

To examine the effect of FET protein binding on the transcriptome, changes on transcript levels were monitored upon FET perturbation by next generation paired-end mRNA-Seq. Transcript expression levels were estimated from the sequencing data using the CUFFLINKS RNA-Seq toolchain (Trapnell et al. 2010). Hierarchical clustering showed that the changes in mRNA abundance induced by knockdown using two siRNAs for the same protein are more similar than to those targeting the other proteins (Fig.12).

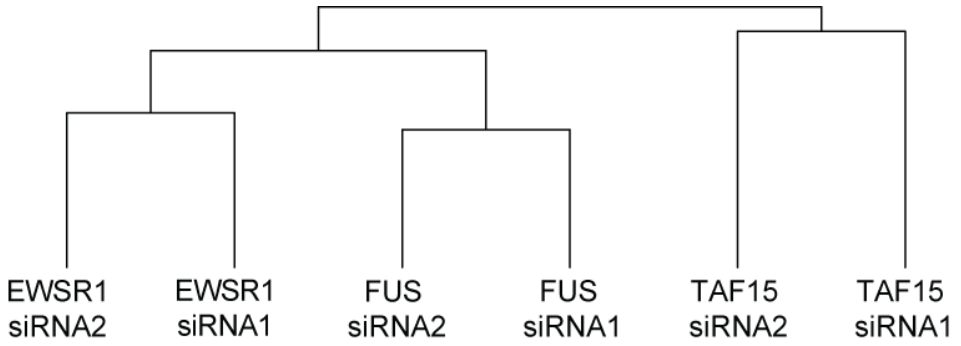


Fig. 12: Hierarchical clustering of mRNA target changes upon FET protein depletion.
Hierarchical clustering diagram of log 2 fold mRNA changes after FUS, EWSR1 and TAF15 knockdown using two different siRNAs for the same protein.

To compare the effect of depletion of FUS, EWSR1 or TAF15 on various sets of transcripts, cumulative density fractions of changes in mRNA abundance were calculated (Fig.13A). For FUS, EWSR1 and TAF15 mRNA abundance of the identified mRNA targets by PAR-CLIP are significantly more downregulated than non-targets (black), confirming the overall functionality of PAR-CLIP targets. The top 1000 mRNA targets with most binding sites showed strongest downregulation.

In addition, downregulation of selected mRNA targets upon FUS knockdown were validated by qRT-PCR (Fig.13B). Since most of the binding sites were observed in intronic sequences, these observations suggest that FET proteins facilitate the processing of their target transcripts.

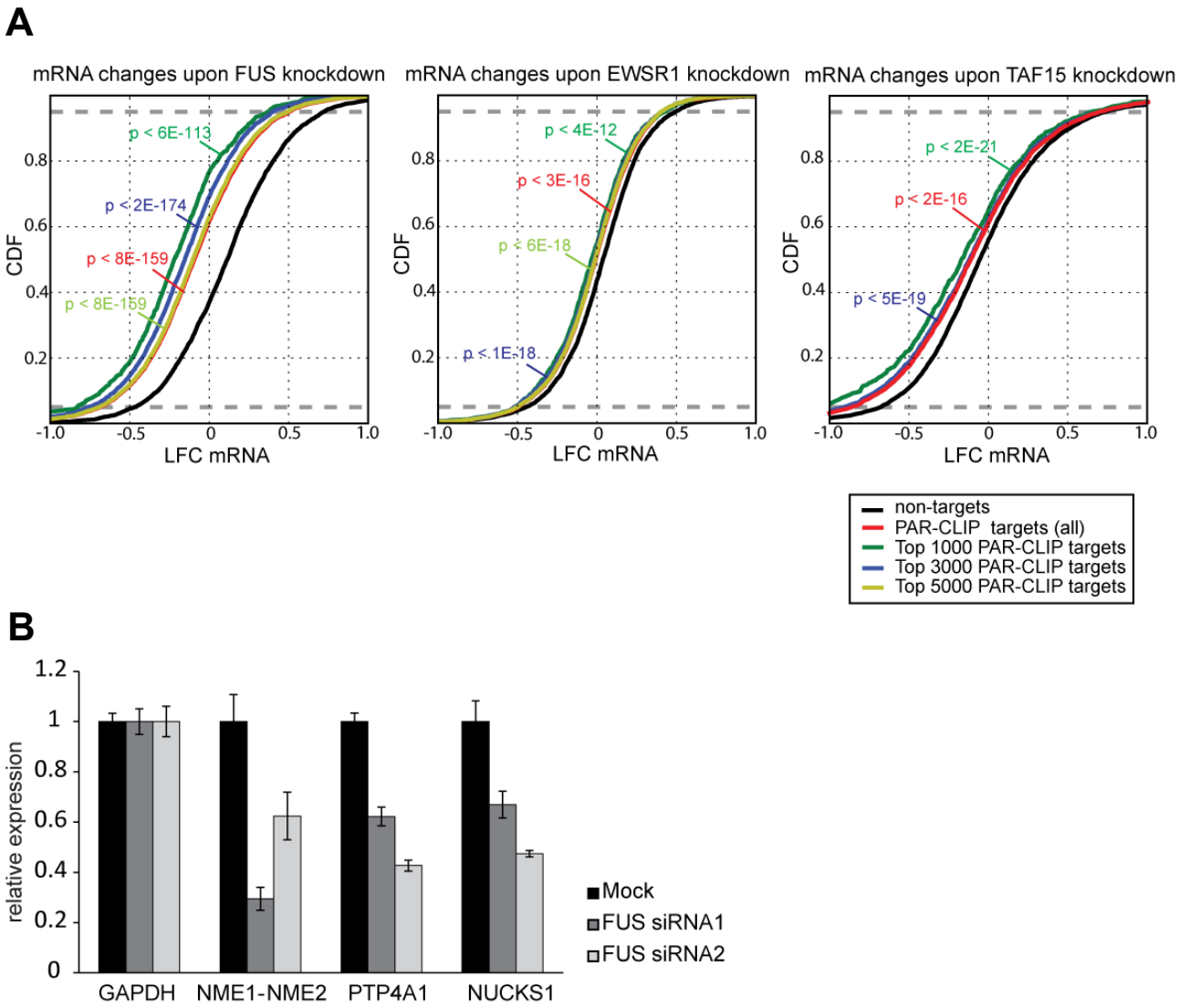


Fig. 13: Effects on mRNA expression of target transcripts after FET protein knockdown.

(**A**) Cumulative distribution functions of mRNA log 2 fold changes. Abundance of FET target mRNAs are downregulated upon knockdown of FUS, EWSR1 and TAF15 compared to non-targeted mRNAs. The p-values indicate the significance of difference between the changes of target versus nontarget transcripts, as given by Wilcoxon rank-sum test. (**B**) Validation of downregulation of selected targets after FUS knockdown by quantitative RT-PCR. Expression of target mRNAs was normalized to GAPDH and is shown relative to mock transfection. Error bars represent the standard deviation for three technical replicates.

3.7 FUS and EWSR1 dependent changes in alternative splicing

To assess the distribution of FET protein binding within introns especially around splice sites, the neighborhood of such sites was scanned for the presence or absence of PAR-CLIP reads. FUS, EWSR1, and TAF15 bind to proximal intronic sequences with a preference toward the 5' and 3' splice sites (Fig.14). For all three FET proteins a sharp peak could be seen around 20 nucleotides upstream of the 3' splice site and for FUS and EWSR1 also a broader peak downstream of the 5'splice site. Similarly, the intronic binding sites of TARDBP were found close to splice sites, as demonstrated previously (Tollervey et al. 2011). In contrast, compared to the wild type FUS protein, the FUS mutant R495X shows reduced binding upstream of the 3' splice site and downstream of the 5' splice site. The stronger binding pattern of all three FET proteins upstream of the 3' splice site is overlapping with the predicted polypyrimidine tract which is located 5 to 40 base pairs before the 3' end of the intron (Reed 1989). This confirms the previous observation that the FET proteins might be involved in regulation of alternative splicing as previously published (Hallier et al. 1998; Paronetto et al. 2011; Ibrahim et al. 2013).

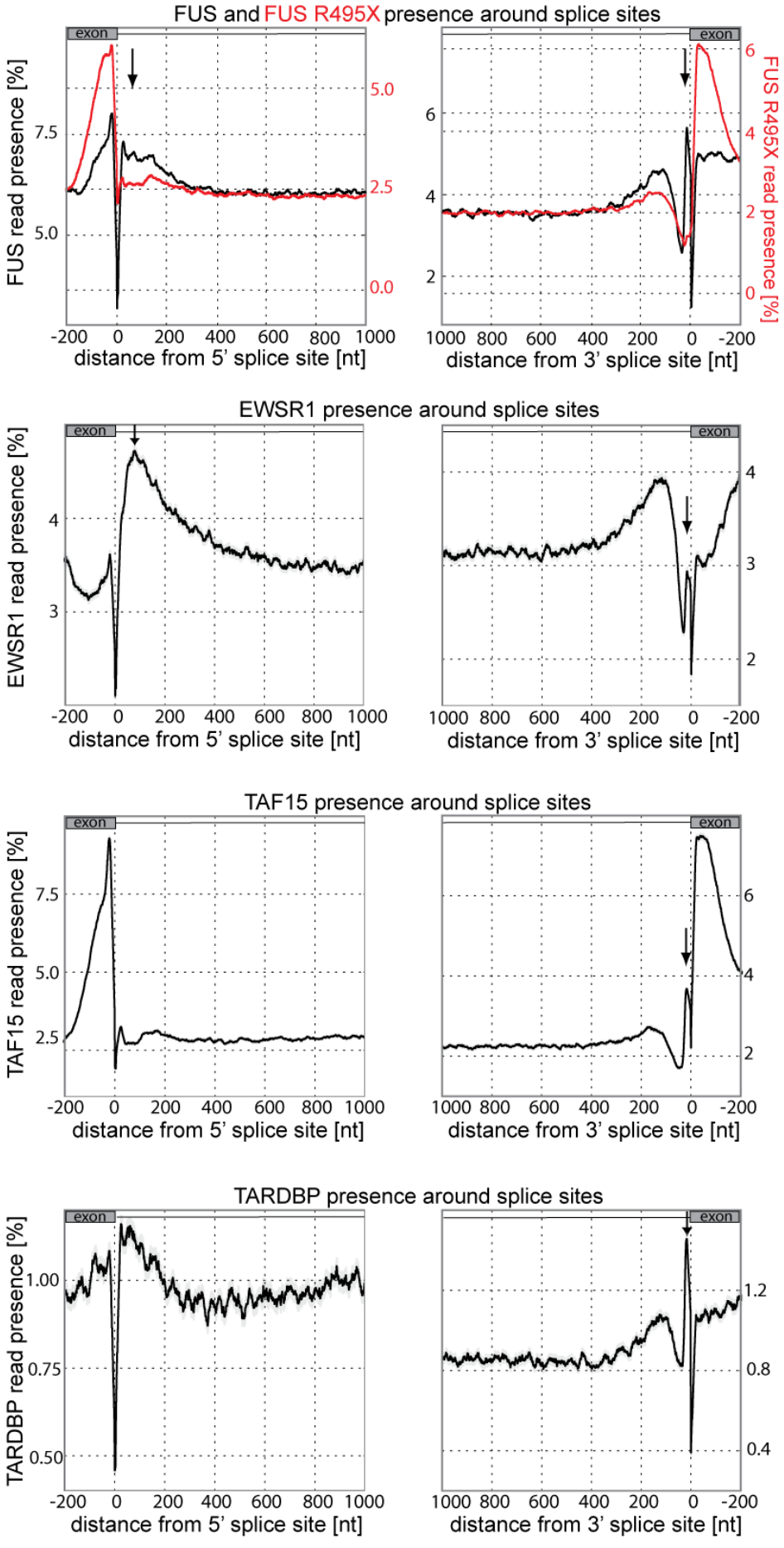

FUS and FUS R495X presence around splice sites
exon
FUS read presence [%]
FUS R495X read presence [%]
distance from 5' splice site [nt]
distance from 3' splice site [nt]
EWSR1 presence around splice sites
EWSR1 read presence [%]
TAF15 presence around splice sites
TAF15 read presence [%]
TARDBP presence around splice sites
TARDBP read presence [%]

Fig. 14: Distribution of FET binding sites around splice sites.
The vicinity of 5' and 3' splice sites of all RefSeq transcripts were scanned for the presence of at least one PAR-CLIP read of FUS, FUS R495X, EWSR1, TAF15 and TARDBP (PAR CLIP read density on a relative scale from 0 to 100%). FET and TARDBP PAR-CLIP reads are enriched near 3' and 5'splice sites (arrows) in contrast to the FUS mutant R495X.

The binding of FET proteins to proximal intronic sequences, the association with splicing factors (Meissner et al. 2003) and already published experimental evidence indicate an involvement of FUS and EWSR1 in alternative splicing. FUS was shown to regulate alternative splicing in mouse neurons pointing to a connection of RNA processing and the development of ALS (Kino et al. 2011; Rogelj et al. 2012). EWSR1 depletion in HeLa cells leads to an altered splicing of genes involved in DNA repair and genotoxic stress responses (Paronetto et al. 2011). This prompted us to examine the mRNA sequencing data for genes with FUS, EWSR1, and TAF15-dependent splicing. Therefore, the RNA-Seq data after FET knockdown were screened for differences in expression of alternative spliced isoforms using the quantification utility CUFFDIFF from the CUFFLINKS RNA-Seq toolchain (Trapnell et al. 2010). For FUS the PSI (Percent Spliced In) ratio was quantified for 13 candidate exons with PCR amplification of both isoforms and Bioanalyzer analysis of the PCR products. PSI (percent spliced in) values were calculated as the molar ratio of the peak corresponding to the exon containing isoform and the sum of the peaks representing both isoforms. For FUS splicing changes for 7 out of 13 in two independent biological replicates could be validated (Fig.15). For EWSR1 the splicing change for 3 out of 7 in two independent biological replicates was validated (Fig.15).

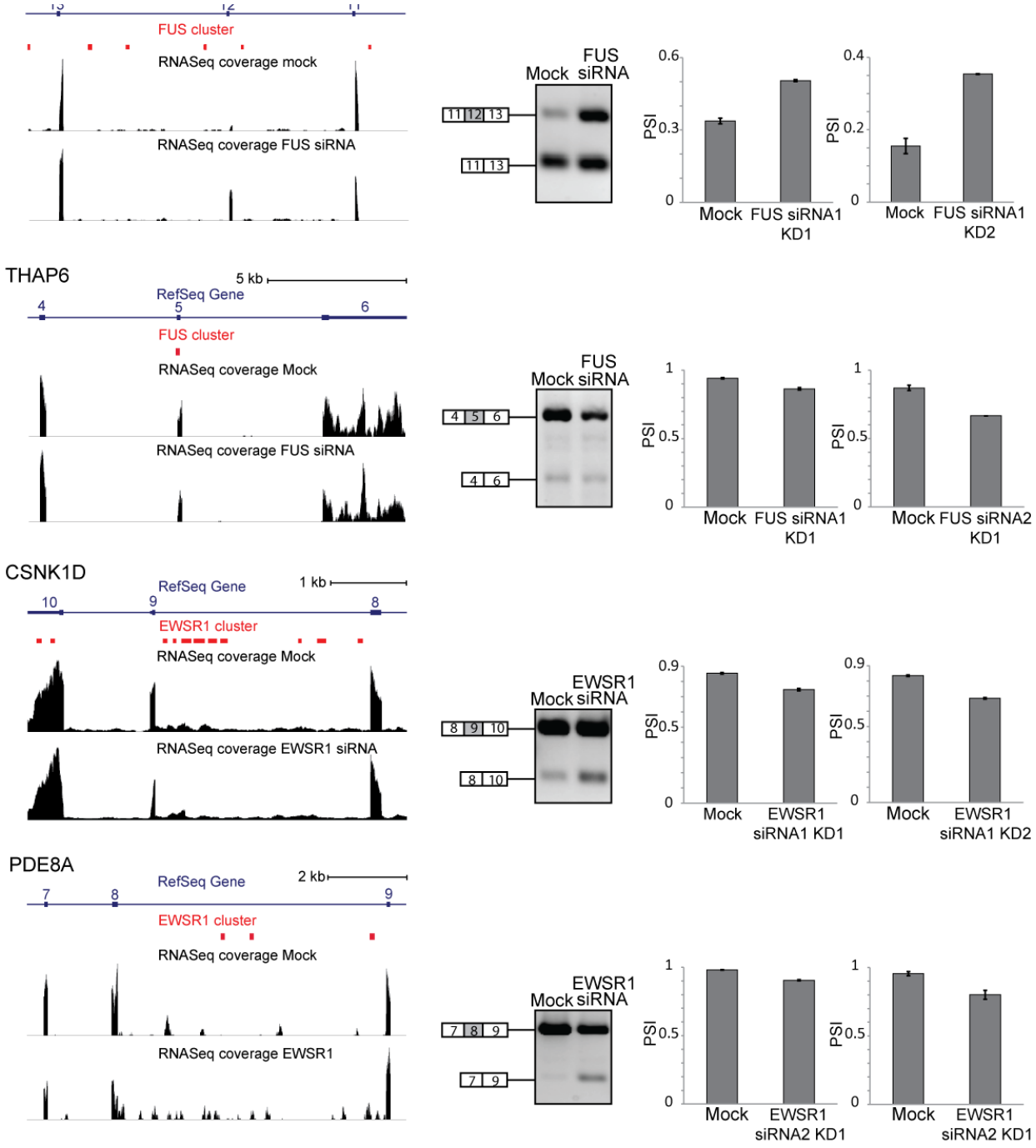


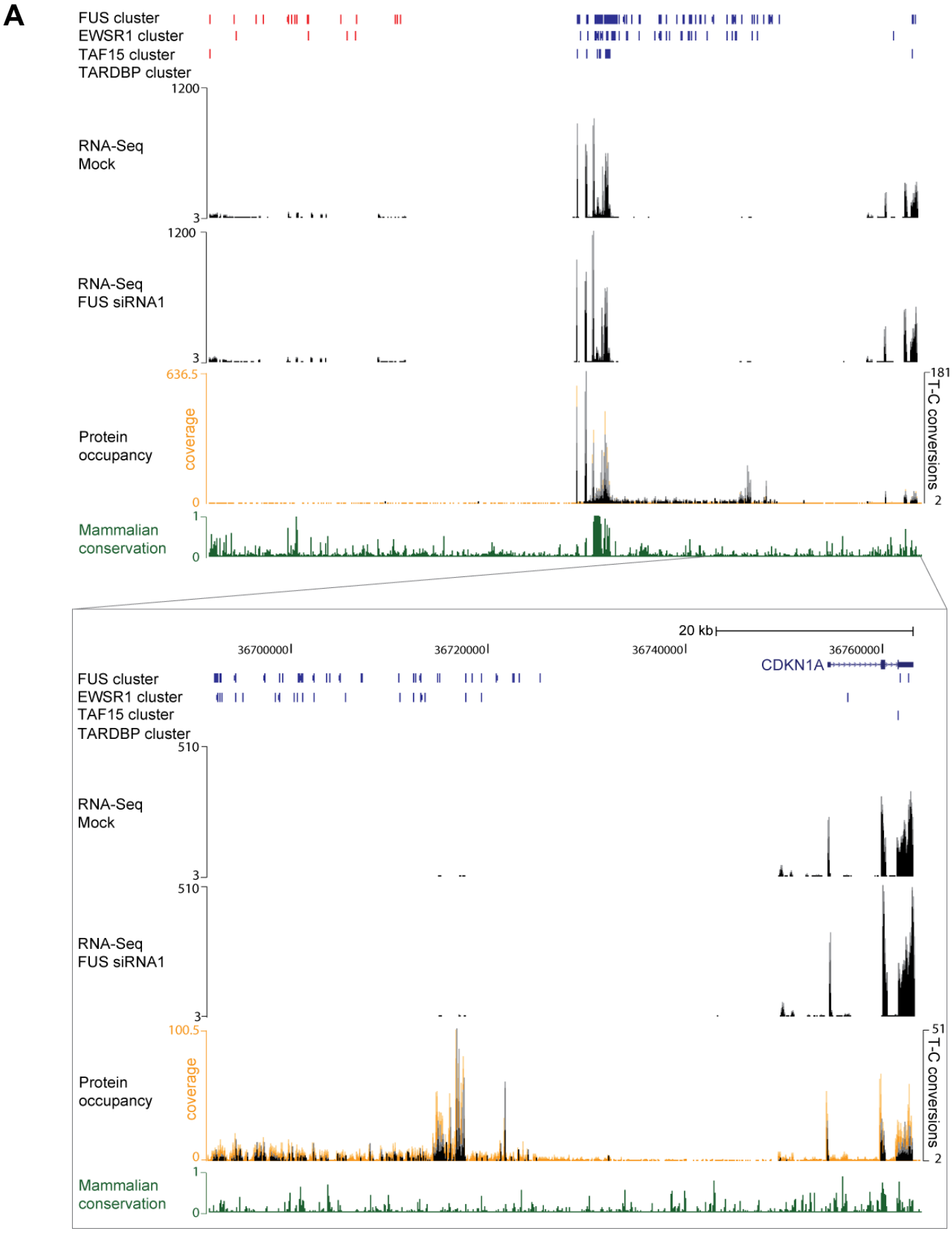
Fig. 15: Effects on alternative spliced exons after FUS and EWSR1 depletion.

Genomic loci centered on alternative exons flanked by constitutive exons are shown in blue. FUS and EWSR1 binding sites (red) are shown together with coverage profiles of RNA-Seq for mock- and siRNA-transfected cells. Differential inclusion of alternative exons was validated by PCR with primers to the flanking exons. The PSI (Percent Spliced In) value was calculated from molar ratios of the PCR products quantified by Bioanalyzer. Error bars represent the standard deviation for three technical replicates.

3.8 Knockdown of FET protein alters mRNA expression of the *SRSF3* locus

Since FET proteins were implicated in transcriptional regulation by associating with RNA Pol II and the transcription factor TFIID (Bertolotti et al. 1996), a focused investigation of FET proteins in transcription based on the example of the *SRSF3* region was performed. For FUS, EWSR1 and TAF15 binding sites were found along the *SRSF3* gene and for FUS and EWSR1 also around 50 kb further downstream (Fig.16A). Furthermore transcription of this locus is supported by RNA-Seq coverage of mock and FUS siRNA-transfected cells. RT-PCR revealed that this transcript likely is a read-through product of the serine/arginine-rich factor 3 (*SRSF3*) gene. Furthermore, this observation is supported by protein occupancy profiling on mRNA (Baltz et al. 2012). For the *SRSF3* locus, but also 50 kb downstream, high levels of T-C transitions were observed indicative for regions of increased protein binding. SRSF3 belongs to the serine/arginine (SR)-rich family which is part of the spliceosome. SR proteins have been shown to be critical for mRNA splicing as well as mRNA export (Huang and Steitz 2001). Furthermore, SRSF3 is critical for cell proliferation and tumor induction because knockdown of SRSF3 leads to cell cycle arrest and apoptosis in U2OS sarcoma cells (Jia et al. 2010). In addition, SRSF3 was identified as an important internal ribosome entry site (IRES) trans-acting factor for poliovirus IRES-mediated translation (Fitzgerald and Semler 2011). Upon poliovirus infection SRSF3 re-localizes from the nucleus to cytoplasm where it interacts with PCBP2 to recruit ribosomes.

Quantitative RT-PCR of FET-depleted cells revealed no change in SRSF3 mRNA level but a 2.5-fold increase of RNA derived from a locus 50 kb downstream of *SRSF3* (Fig.16B). Furthermore a 4-fold upregulation in expression of the *SRSF3* flanking gene *CDKN1A* and mild upregulation of *STK38* was seen. These data raised the possibility that FUS and EWSR1 might influence transcription of this genomic region. Another study by Schwartz et al., which was published meanwhile, showed that loss of FUS leads to an accumulation of RNA Pol II at the transcriptional start site (TSS) and premature polyadenylation of transcripts (Schwartz et al. 2012). In their study, FUS knockdown led to production of shorter isoforms for 47 manually checked mRNAs and longer isoforms for 15 mRNAs. In our study, FET protein knockdown for the *SRSF3* locus led to production of a longer *SRSF3* mRNA isoform.


A
FUS cluster
EWSR1 cluster
TAF15 cluster
TARDBP cluster
1200
RNA-Seq
Mock
3
1200
RNA-Seq
FUS siRNA1
3
636.5
Protein
occupancy
coverage
0
181
T-C conversions
2
1
Mammalian
conservation
0
20 kb
36700000
36720000
36740000
36760000
CDKN1A
FUS cluster
EWSR1 cluster
TAF15 cluster
TARDBP cluster
510
RNA-Seq
Mock
3
510
RNA-Seq
FUS siRNA1
3
100.5
Protein
occupancy
coverage
0
51
T-C conversions
2
1
Mammalian
conservation
0

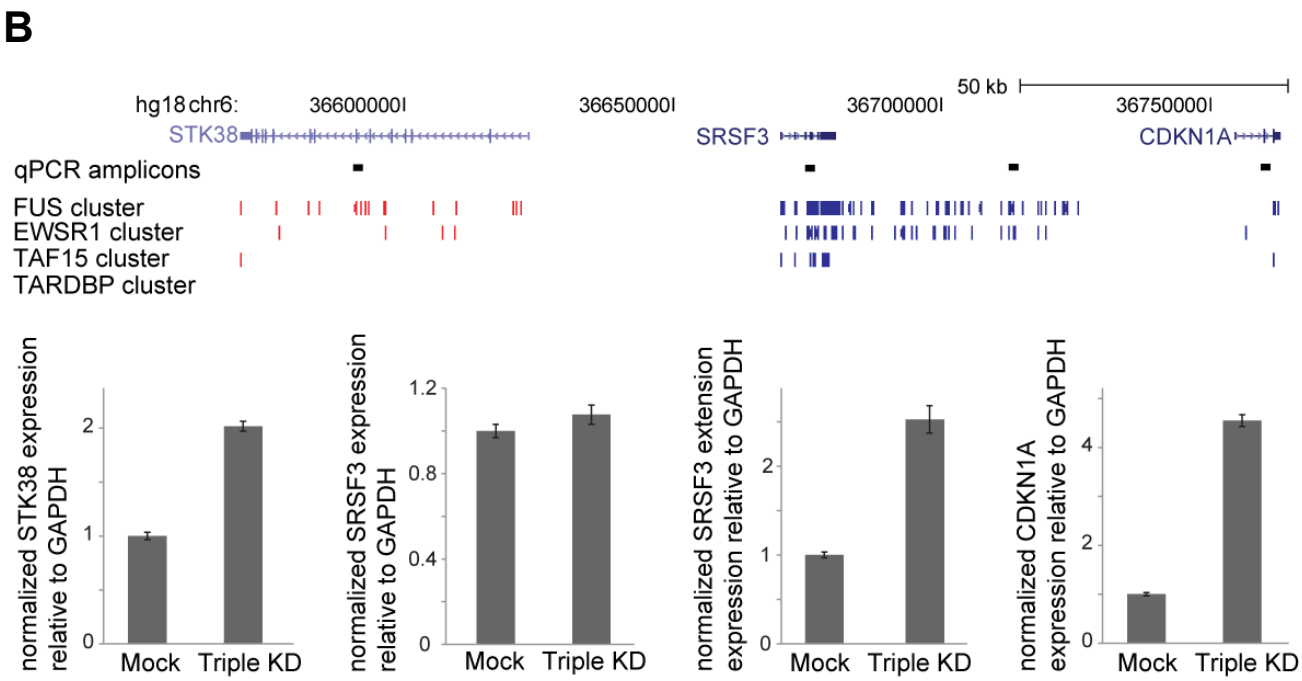


Fig. 16: Specific binding of FUS and EWSR1 to a downstream region of *SRFS3*.

(**A**) Distribution of binding sites of FET proteins and TARDBP around the *SFRS3* locus. RNA-Seq coverage of FUS siRNA1 and mock transfected cells, as well as protein occupancy (orange) together with T-C conversions (black) and the PhyloP mammalian conservation score (green) are also shown. (**B**) Quantitative RT-PCR results are shown after depletion of all three FET proteins at the same time (Triple KD) for *STK38*, *SRFS3*, a region 50 kb downstream of *SRFS3* (*SRSF3* extension) and *CDKN1A* are shown together with the respective location of the qPCR amplicons. Expression levels were normalized to GAPDH. Error bars represent the standard deviation for three technical replicates.

Next, we asked whether this is a direct effect of the FET proteins on transcription by interaction with RNA Pol II. A role for the FET proteins in transcription was already suggested by several studies. All three FET proteins were found to co-purify with the transcription factor TFIID and RNA Pol II (Bertolotti et al. 1996; Hoffmann and Roeder 1996; Bertolotti et al. 1998). Furthermore, FUS interacts with the TATA-binding protein which enhances RNA Pol II transcription (Tan and Manley 2010). Schwartz and coworkers revealed that FUS is able to bind the C-terminal domain (CTD) of RNA Pol II orchestrating CTD phosphorylation (Schwartz et al. 2012). Transcriptional elongation is influenced by the phosphorylation state of the CTD which serves as platform for factors regulating transcription. Phosphorylation activates RNA Pol II and triggers transcriptional initiation. Phosphorylation occurs mainly at Ser2 and Ser5 of the heptapeptide repeat. The phosphorylation status of the CTD changes at different positions along the transcription unit. Phosphorylation of Ser5 residues predominates near the beginning of genes, whereas polymerases near the ends of genes are extensively phosphorylated on Ser2 residues (Komarnitsky et al. 2000; Morris et al. 2005).

Schwartz and colleagues suggested that FUS modulates the amount of Ser2 phosphorylation around the TSSs as loss of FUS leads to an accumulation of phosphorylated Ser2 at the TSSs (Schwartz et al. 2012).

To test an influence of FET proteins on transcription of the *SRSF3* locus, genome-wide localization of RNA Pol II on the chromatin of HEK293 cells using ChIP followed by qRT-PCR was examined. Three different RNA Pol II antibodies were used to immunoprecipitate different phosphorylated forms of RNA Pol II. The RNA Pol II antibody CTD4H8 (ab5408, Abcam) binds both unphosphorylated and Ser2- and Ser5 –phosphorylated forms of the heptapeptide repeat YSPTSPS within the CTD of RNA Pol II. Using antibodies specifically recognizing phosphorylated Ser2 or Ser5 of RNA Pol II for ChIP experiments it is feasible to distinguish between initiating and terminating polymerase. To test whether the FET proteins have an influence on transcription, FET proteins were depleted and ChIP followed by qRT-PCR was performed. Cells were treated with FUS, EWSR1 and TAF15 siRNAs to knockdown expression of all three FET proteins (triple knockdown) or left untreated (mock). After ChIP the immunoprecipitated DNA was analyzed by qRT-PCR using primer specific for the *SRSF3* locus and further downstream (Fig.17).

Triple knockdown of FET proteins resulted in an increase of RNA Pol II at the TSS of *SRSF3* simultaneously with a slight increase of Ser2- and Ser5-phosphorylated RNA Pol II forms. At the end of the *SRSF3* gene and further downstream the amount of polymerase decreases whereas Ser2- and Ser5- phosphorylated forms remain nearly similar after knockdown. Consistently, Schwartz et al. also showed an increase of total RNA Pol II near TSSs but in contrast to our data they also showed a significant increase of Ser2 phosphorylation around TSSs. This indicates that loss of the FET proteins leads to an accumulation of RNA Pol II at the TSS and a decrease at the end of the gene. In the case of *SRSF3*, changes in the phosphorylation state of RNA Pol II CTD do not seem to be the reason for the altered expression of the longer SRSF3 mRNA isoform as verified by qRT-PCR after FET protein depletion.

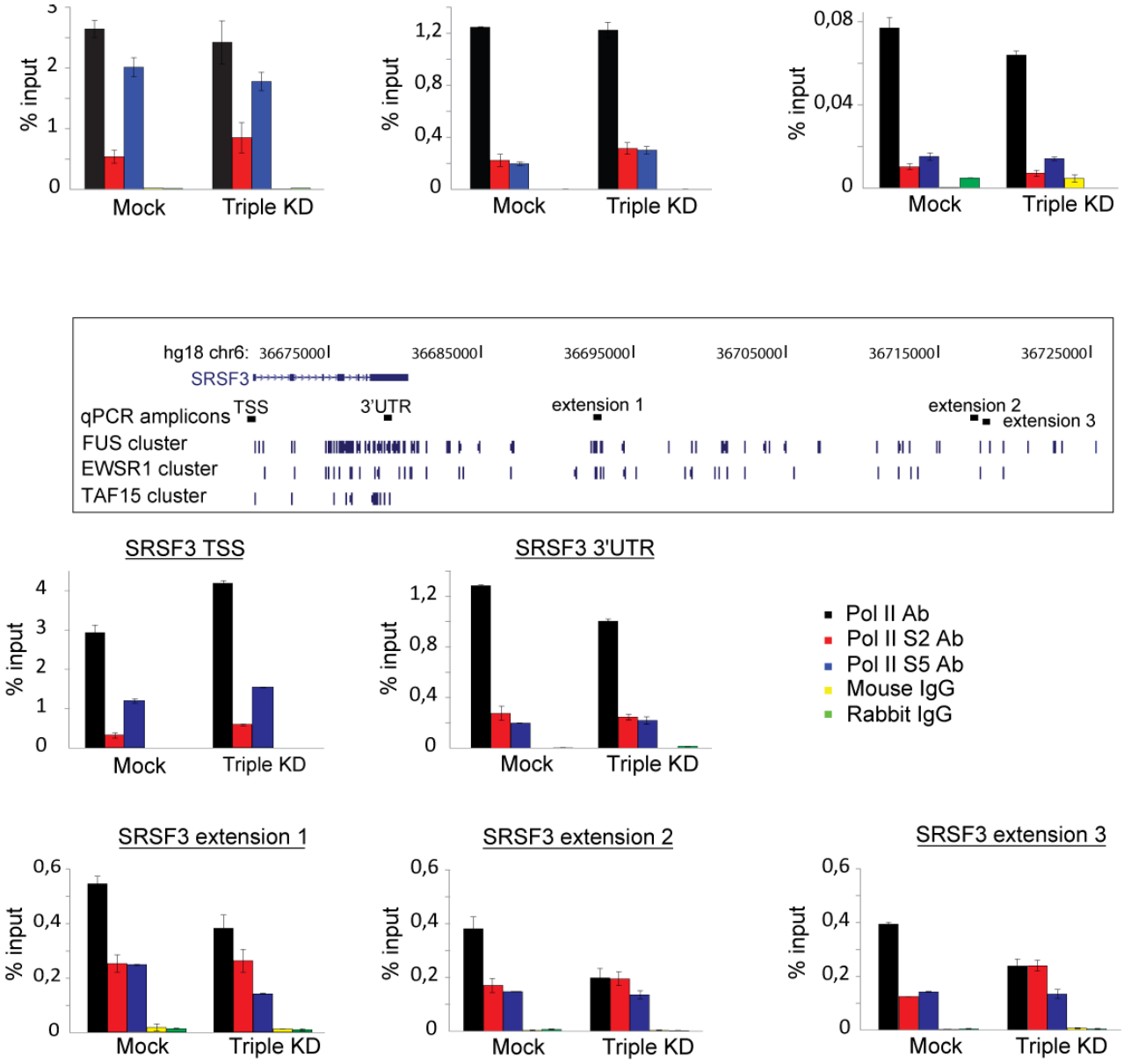


Fig. 17: Loss of FET proteins alters distribution of RNA polymerase II around *SRSF3*.

Various regions around the *SRSF3* gene were tested for RNA Pol II occupancy using ChIP. Antibodies recognizing both unphosphorylated and phosphorylated Ser5 residues of the RNA Pol II (Pol II Ab), recognizing only phosphorylated Ser2 (Pol II S2 Ab) and recognizing only phosphorylated Ser5 residues (Pol II S5 Ab) were used for immunoprecipitation. Mouse IgG and rabbit IgG were used as negative controls for unspecific binding. The immunoprecipitated DNA was quantified by qRT-PCR. Error bars represent the standard deviation for three technical replicates.

3.9 FET proteins bind antisense, non-coding RNAs at promoter regions

Since RBPs can interact with different classes of RNAs (Cabili et al. 2011), binding of the FET proteins to long intergenic non-coding RNAs (lincRNAs) was examined as well. About 1% of the FUS and EWSR1 and 0.5% of the TAF15 sequence clusters mapped to ncRNAs, supporting previous findings that the FET proteins interact with the promoter ncRNA$_{CCND1}$ which negatively regulates transcription of CCND1 *in cis* by forming a ribonucleoprotein repressor complex (Wang et al. 2008). To investigate whether FET proteins have a more widespread function in the regulation of cell growth by associating with ncRNAs derived from promoters of cell cycle regulators, 216 transcribed regions that encode putative ncRNAs within promoters of 49 human cell-cycle genes (cyclins, cyclin-dependent kinases and cyclin-dependent kinase inhibitors) were scanned for possible binding sites. In 10 of the 49 genetic loci FUS, EWSR1 and TAF15 sequence clusters mapped to ncRNAs (Tab.3). Interaction with this subset of promoter ncRNA seemed to be specific to FET proteins. TARDBP, although described to bind to several ncRNAs (Tollervey et al. 2011), and the FUS variant R495X, yielded no sequence clusters mapping to these genomic regions (Fig.18). In particular for FUS it has been already shown that it is often bound to RNA antisense upstream of promoter regions leading to a down regulation of transcription of the sense strand (Ishigaki et al. 2012). EWSR1 and TAF15 PAR-CLIP data revealed RNA binding sites antisense to promoters of genes involved in cell cycle regulation (Fig.18). These findings suggest that FUS, EWSR1 and TAF15 might have a specific and more pervasive function, mediated by promoter ncRNAs, in transcriptional regulation of cell cycle genes.

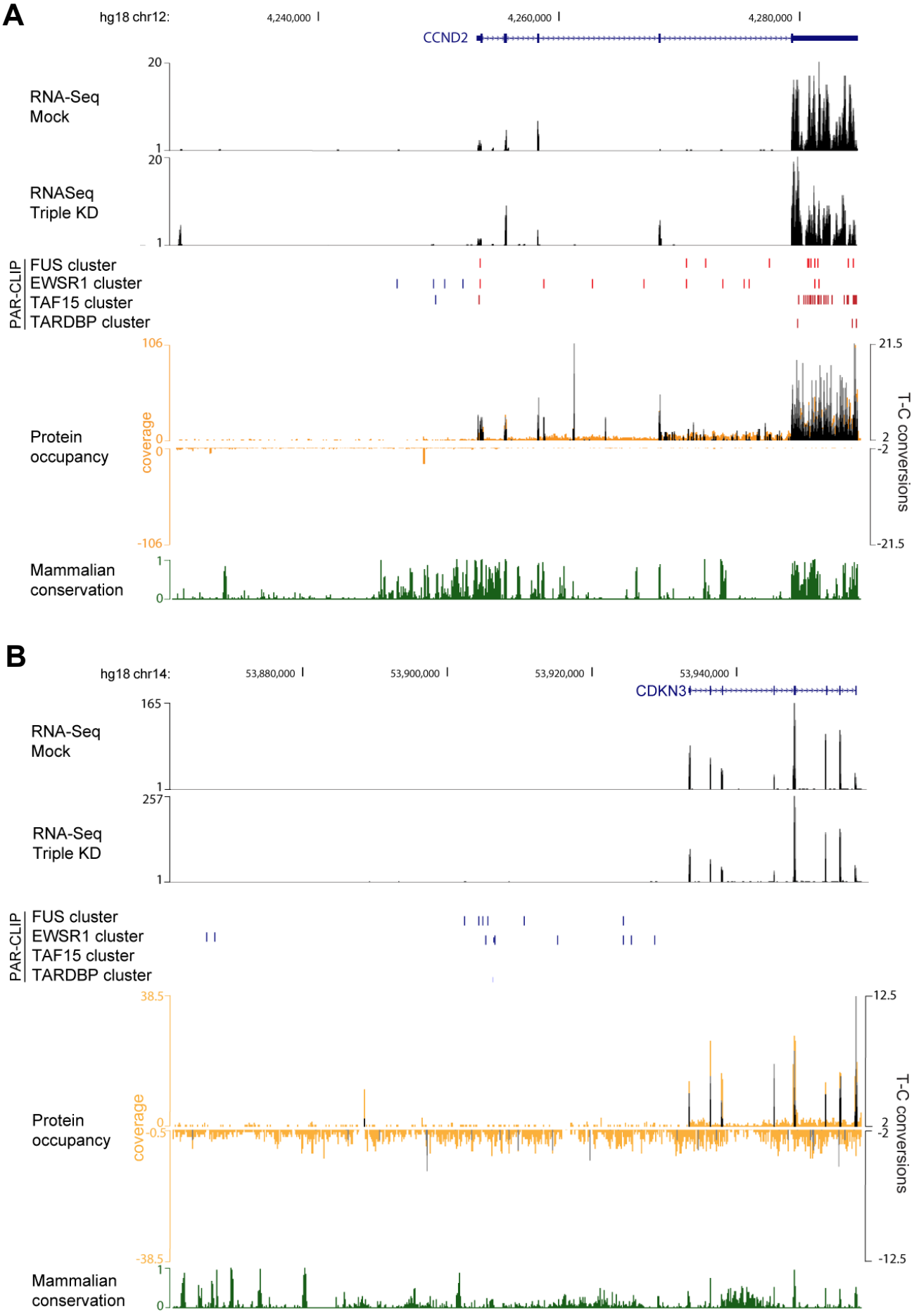

A
hg18 chr12:
4,240,000
4,260,000
4,280,000
CCND2
RNA-Seq
Mock
RNASeq
Triple KD
PAR-CLIP
FUS cluster
EWSR1 cluster
TAF15 cluster
TARDBP cluster
Protein
occupancy
coverage
T-C conversions
Mammalian
conservation
B
hg18 chr14:
53,880,000
53,900,000
53,920,000
53,940,000
CDKN3
RNA-Seq
Mock
RNA-Seq
Triple KD
PAR-CLIP
FUS cluster
EWSR1 cluster
TAF15 cluster
TARDBP cluster
Protein
occupancy
coverage
T-C conversions
Mammalian
conservation

Fig. 18: Representative examples of FET protein binding to intergenic regions in the proximity of promoter regions.

RNA-Seq coverage of FUS, EWSR1 and TAF15 siRNA transfected cells (Triple KD) and untransfected cells (Mock) is shown together with FUS, EWSR1, TAF15 and TARDBP sense (blue) and antisense binding sites (red) determined by PAR-CLIP around *CCND2* (**A**) and *CDKN3* (**B).** Furthermore protein occupancy profiling data on mRNA (Baltz et al. 2012) with sequence coverage in orange and T-C transition profile in black is given together with the UCSC PhastCons conservation of placental mammals in green.

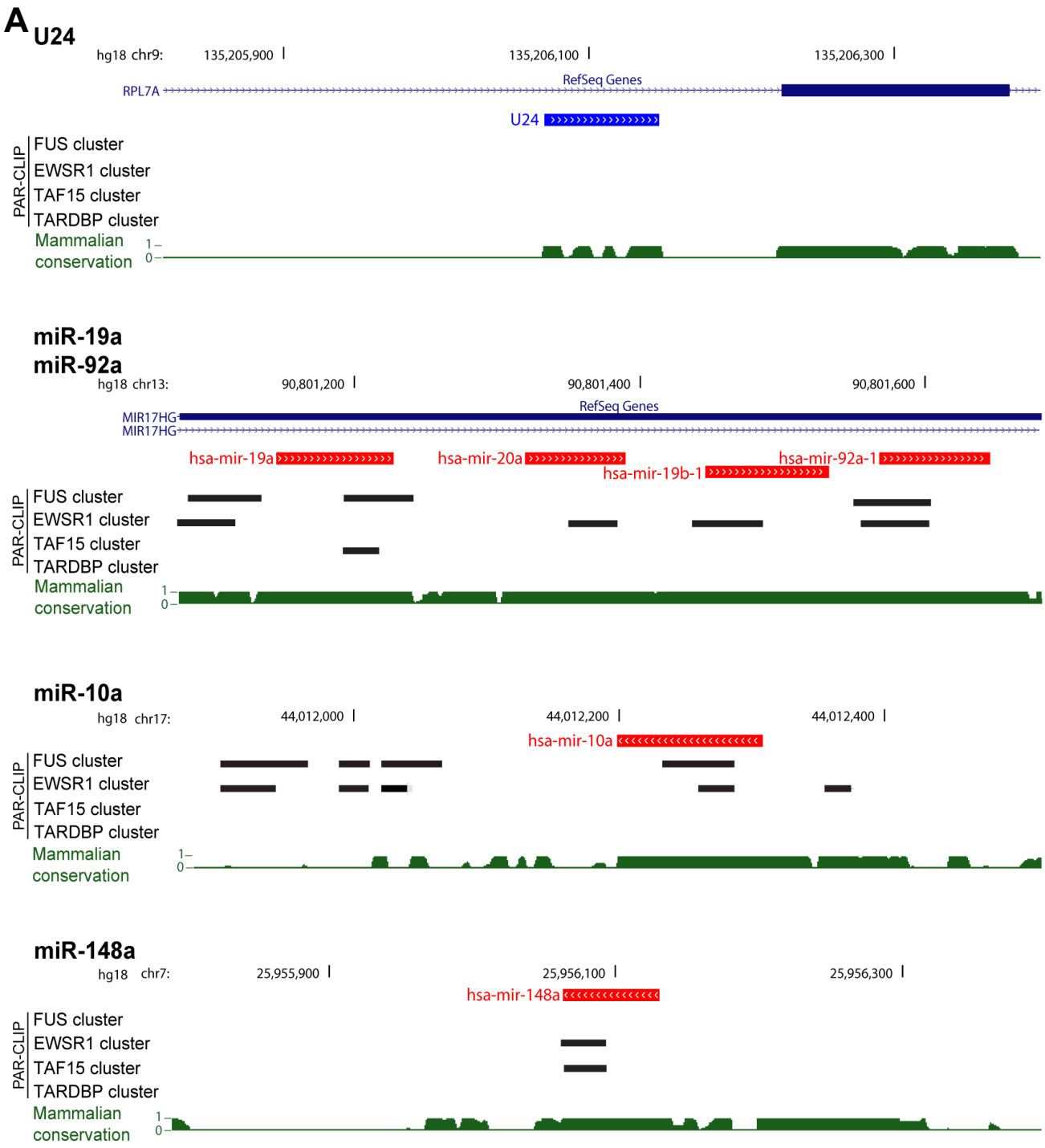
Tab. 3: FET binding sites at transcripts derived from promoter regions of cell cycle genes

Gene name	Chromosome	FUS cluster	EWSR1 cluster	TAF15 cluster
CDKL5	chrX	-	-	-
CCNB3	chrX	-	-	-
CDKN2A	chr9	-	-	-
ARF	chr9	-	+	-
CDKN2B	chr9	+	-	-
CDK9	chr9	-	-	-
CCNE2	chr8	-	+	-
CDK6	chr7	-	-	-
CDK5	chr7	-	+	-
CDKN1A	chr6	-	+	-
CCND3	chr6	-	-	-
CCNC	chr6	-	+	-
CCNO	chr5	-	-	-
CDKL3	chr5	-	-	-
CCNJL	chr5	-	-	-
CCNG1	chr5	-	-	-
CCNI	chr4	-	-	-
CCNG2	chr4	-	-	-
CCNA2	chr4	-	-	-
CCNL1	chr3	-	-	-
CDKL4	chr2	-	-	-
CNNM4	chr2	-	-	-
CNNM3	chr2	-	-	-
CCNT2	chr2	-	-	-
CCNYL1	chr2	-	-	-
CDK5R2	chr2	-	-	-
CDKN2D	chr19	-	-	-
CCNE1	chr19	-	-	-
CDK5R1	chr17	+	+	+
CDK3	chr17	-	-	-
CCNF	chr16	-	-	-
CDK3	chr17	-	-	-
CCNF	chr16	-	-	-
CCNYL3	chr16	-	-	-
CDK10	chr16	-	-	-
CCNB2	chr15	-	-	-
CDKN3	chr14	-	-	-
CCNK	chr14	-	+	-
CDK8	chr13	-	-	-
CCNA1	chr13	-	-	-
CCND2	chr12	-	+	+
CDKN1B	chr12	-	-	-
CCNT1	chr12	-	-	-
CDK2	chr12	-	-	-
CDK4	chr12	-	-	-
CDKN1C	chr11	-	-	-
CCND1	chr11	-	-	-
CCNYL2	chr10	-	-	-
CNNM1	chr10	-	-	-
CCNL2	chr1	-	-	-
CDKN2C	chr1	+	-	-

List of cell cycle genes whith putative ncRNA expression at their promoters and presence (+) or absence (-) of FET binding sites.

3.10 FET proteins modulate miRNA expression

RBPs are key components in the determination of miRNA function as they control different stages of miRNA biogenesis, localization and activity. In particular, FET proteins were implicated to be involved in miRNA biogenesis based on their presence in the microprocessor, the Drosha-containing protein complex catalyzing the first step in miRNA maturation (Gregory et al. 2004). Furthermore, FUS contributes to biogenesis of a subset of miRNAs involved in neuronal function and differentiation (Morlando et al. 2012). Also TAF15 regulates gene expression of cell cycle regulatory genes post-transcriptional through a pathway involving miRNAs (Ballarino et al. 2012).
Since FET protein sequence clusters mapped to several pre-miRNA precursors (Fig.19A), the effect of FUS, EWSR1 and TAF15 on miRNA expression was investigated by sequencing small RNA libraries from mock- and siRNA-transfected cells. A comparison of the small RNA libraries revealed only mild changes in miRNA expression upon RNAi-mediated protein knockdown for the single FET proteins (Supplementary Tab.1). Quantitative RT-PCR on a selected number of miRNAs confirmed these findings (Fig.19B). Upon FUS knockdown expression of miR-10a is decreased whereas expression of miR-19a, miR-92a and miR-148 does not change. Interestingly, triple knockdown of all three FET proteins at the same time lead to downregulation of miR-19a, miR-92a, miR-10a and miR148a. This indicates some degree of redundancy of FET proteins.


A
U24
hg18 chr9:
135,205,900
135,206,100
135,206,300
RefSeq Genes
RPL7A
U24
PAR-CLIP
FUS cluster
EWSR1 cluster
TAF15 cluster
TARDBP cluster
Mammalian conservation
1
0
miR-19a
miR-92a
hg18 chr13:
90,801,200
90,801,400
90,801,600
RefSeq Genes
MIR17HG
MIR17HG
hsa-mir-19a
hsa-mir-20a
hsa-mir-92a-1
hsa-mir-19b-1
PAR-CLIP
FUS cluster
EWSR1 cluster
TAF15 cluster
TARDBP cluster
Mammalian conservation
1
0
miR-10a
hg18 chr17:
44,012,000
44,012,200
44,012,400
hsa-mir-10a
PAR-CLIP
FUS cluster
EWSR1 cluster
TAF15 cluster
TARDBP cluster
Mammalian conservation
1
0
miR-148a
hg18 chr7:
25,955,900
25,956,100
25,956,300
hsa-mir-148a
PAR-CLIP
FUS cluster
EWSR1 cluster
TAF15 cluster
TARDBP cluster
Mammalian conservation
1
0

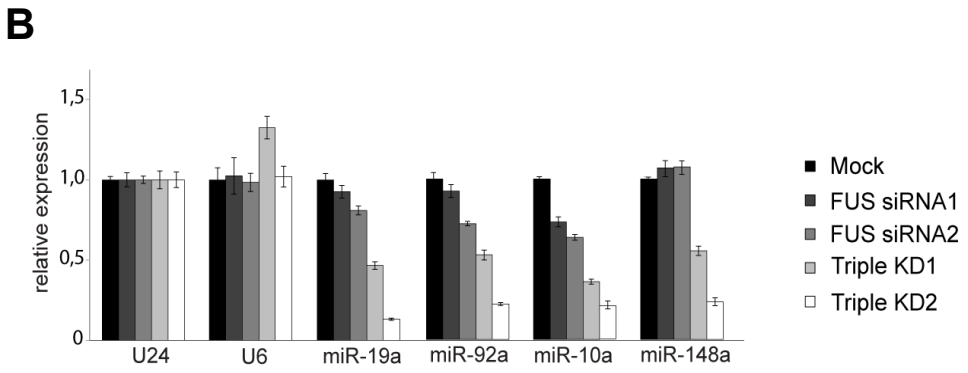

B
relative expression
1,5
1,0
0,5
0
U24
U6
miR-19a
miR-92a
miR-10a
miR-148a
Mock
FUS siRNA1
FUS siRNA2
Triple KD1
Triple KD2

Fig. 19: Effects on miRNA expression after depletion of FET proteins.

(A) Distribution of FET clusters around the U24 locus and several miRNA cluster. **(B)** Upon FUS knockdown, miRNA abundance did not change. Total RNA of siRNA- and mock-transfected cells was subjected to TaqMan qRT-PCR and expression of U6 (control), miR-19a, miR-92a, miR10a and miR148a was normalized to the expression of the endogenous control U24 snRNA.

4. Discussion

RBPs are playing an important role in posttranscriptional gene regulation. However, the function of many RBPs is still unknown. One family of such RBPs constitute FUS/TLS, EWSR1 and TAF15 which are multifunctional proteins involved in several cancer types like sarcomas, leukemias but also in neurodegenerative diseases including ALS and FTLD. Nevertheless, despite emerging information of FET protein functions many aspect of their involvement in gene regulation are elusive. The main goal of this study was to elucidate the role of the FET family members in transcriptional and posttranscriptional gene regulation.

4.1 FET proteins bind mainly intronic sequences

To determine the transcriptome wide binding sites for FUS, EWSR1 and TAF15 at a nucleotide resolution PAR-CLIP was used. FET proteins bind mostly to introns, consistent with the nuclear localization of the proteins (Andersson et al. 2008), but also to 3'UTRs and CDS (Fig.7). For TAF15, besides intronic binding also a larger set of binding sites were located in 3'UTRs. This supports the idea that these proteins might still bind to different targets or have different binding sites within the same target, respectively, despite the high identity in their RNA binding domain. However, intersection of the top 1000 targets ranked by T-to-C frequency revealed a large overlap of targets in general (Fig.9). This in accordance with the study by Hoell and coworkers where it was shown that also the binding sites within the RNA targets largely overlap between all three FET family members (Hoell et al. 2011).

For FUS and TAF15 a clear preference to UAC rich regions was seen, whereas EWSR1 seem to bind more GC-rich sequences surprisingly (Fig.10). The different binding motif of EWSR1 indicates binding of different targets or target sites then FUS and TAF15 respectively. This is consistent with the observation that among the top 1000 target RNAs of each FET protein, ranked by the number of binding sites, EWSR1 shows the smallest overlap (Fig.9). This might point to less redundancy between all three FET proteins as initially assumed. Due to the different binding motif of EWSR1 it would be conceivable that FUS and TAF15 might exhibit some redundancy.

Next, binding specificity of all three proteins has to be verified for example by electrophoretic mobility shift assay (EMSA). In contrast to this study, Hoell et al. did not reveal any specific binding motif of FUS, EWSR1 and TAF15 by PAR-CLIP in HEK293 cells (Hoell et al. 2011).

They also reported that secondary structure of the RNA might play an important role in recognition suggesting FUS binding to AU-rich RNA stem loop structures. By contrast, motif analysis of the FET PAR-CLIP data confirmed a preference for binding single-stranded sequences for all three FET proteins (Fig.10). There is no indication of preferential binding to stem loop RNA structures as suggested by Hoell et al.

Comparison with the TARDBP PAR-CLIP data revealed no indication for a similar binding motif of FUS and TARDBP suggesting a difference in target binding.

In general, the large number of RNA binding sites for all three FET proteins implies that these proteins tend to bind RNA frequently which would support a general role of FET proteins in mRNA processing. Non-physiological binding of overexpressed recombinant protein might be also a reason for the large number of binding sites. However, expression levels of recombinant FET proteins were even slightly lower than the expression levels of the endogenous proteins (Fig.5)

The ALS causing FUS mutant R495X showed reduced binding to introns and an enhanced binding to 3' and 5'UTRs presumably due to its altered cytoplasmic localization (Fig.7). It would be conceivable that the FUS mutant bind different targets as a result of the altered localization. This supports the model where ALS is caused by a combination of a loss of function and a gain of toxicity. On the one hand mutant FUS is mainly localized in the cytoplasm and aggregates in ubiquitin-positive inclusions because of a disrupted putative nuclear localization signal at the C-terminus and cannot function fully in the nucleus anymore. On the other hand in the cytoplasm FUS R495X can access different targets leading to a gain of function effect. Both scenarios support the assumption that misregulated RNA processing could trigger neurodegeneration.

However, a recent study by Shelkovnikova et al. indicated that FUS aggregation itself can trigger ALS-like pathology (Shelkovnikova et al. 2013). They could show that a truncated version of FUS with a disrupted NLS and RNA binding motif already aggregates in SH-SY5Y neuroblastoma cells but become excluded from stress granules. However, RNA-binding competent variants are recruited to stress granules. Transgenic mice expressing the RNA-binding incompetent FUS variant develop fatal motor neuron pathology and endogenous FUS is present in pathological inclusions formed by transgenic truncated protein.

Recently, two studies focussed on RNA granule-like structures which are often composed of RBPs containing low-complexity (LC) sequences (Han et al. 2012; Kato et al. 2012). LC sequences are protein regions with little diversity in their amino acid composition, often found in RNA- and DNA-binding proteins (Michelitsch and Weissman 2000). These sequences are also present in the N-terminal end of FUS, EWSR1 and TAF15. The LC sequences of these proteins associated with other RBPs can be reversibly transformed from soluble to polymeric, amyloid-like fibers.

Kato et al. proposed that this transition may help RBPs to move in and out of cellular structures like RNA granules, stress granules and P bodies (Kato et al. 2012). The amyloid-like fibers are able to recruit and retain mRNAs entering neuronal granules similar to RNA granules. Therefore, LC sequences may be responsible for forming pathological aggregates. Interestingly, phosphorylation of the LC domain of FUS prevented retention of these mRNAs (Han et al. 2012). Hence, the dynamics of aggregation and disaggregation might be regulated by posttranslational modifications.
In conclusion, it is feasible that alterations in RNA metabolism may not be the primary cause in ALS but rather be a "second hit" that contributes to irreversible aggregation of FUS.

4.2 FET protein and TARDBP mRNA targets encode proteins involved in protein degradation

Interestingly, analyzing the RNA targets of all three FET proteins as well as TARDBP, ubiquitin-proteasome-related gene categories are overrepresented (Fig.9). There are several lines of evidence that an impairment of protein turnover contribute to the pathogenesis of ALS and other neurodegenerative diseases. Mutations in *UBQLN2*, which encodes the ubiquitin-like protein ubiquilin 2, were shown to cause dominantly inherited ALS (Deng et al. 2011). UBQLN2 is a member of the ubiquitin-like protein family (ubiquilins), which regulates the ubiquitin–proteasome system (UPS) of protein degradation by delivering ubiquitinated proteins to the proteasome. Using an UPS reporter substrate, the investigators could prove a significantly slower degradation of the substrate in cells expressing mutant ubiquilin 2 than in cells expressing the wildtype protein. Furthermore, ubiquilin positive inclusions are common in a wide spectrum of ALS. Whether ubiquilin 2 inclusions initiate neurodegeneration or are a consequence of intracellular deregulation remains to be established.
ALS linked-mutations were also identified in other genes affecting protein homeostasis. Mutations in optineurin (OPTN), a multifunctional protein which is able to bind polyubiquitinated proteins, have been reported in sporadic and familial ALS cases (Maruyama et al. 2010; Del Bo et al. 2011) In addition, mutations in *VCP* were identified in 1-2 % of all familial ALS cases (Johnson et al. 2010). The ATP-driven chaperone valosin-containing protein (VCP) interacts with a wide range of ubiquitinated proteins to enable degradation and regulates critical steps in ubiquitin-dependent protein quality control (Meyer et al. 2012).

Similar to VCP, p62/SQSTM1 has been shown to interact with polyubiquitinated proteins targeting them to the proteasome (Moscat and Diaz-Meco 2012). p62 inclusions have been reported in many neurodegenerative diseases and mutations in p62 have been found in both familial and sporadic ALS patients (Fecto et al. 2011; Brettschneider et al. 2012; Rubino et al. 2012; Teyssou et al. 2013). Although, how all these ALS-associated mutations in OPTN, VCP, p62, TARDBP and FUS contribute to the pathogenesis of ALS has not been established yet, proteasome disturbance seems likely to play an important role.

4.3 FET protein binding leads to increase in target mRNA levels

Based on the RNA-Seq data of FET down-regulated cells, a stabilizing function of target transcripts was observed (Fig.13). siRNA mediated knockdown of FUS, EWSR1 and TAF15 led to highly significant destabilization of transcripts with FET binding sites which confirms the functionality of the targets. The exact mechanisms by which the FET proteins achieve this mRNA stabilization still need to be revealed. Since the majority of FET binding sites lie within introns it would be conceivable that FET proteins play an important role in pre-mRNA processing, especially splicing. If splicing of target mRNAs is disturbed upon FET protein depletion, pre-mRNAs might be degraded.

We also quantified the protein abundance using the SILAC approach (Ong et al. 2002) and compared target protein levels in mock transfected cells versus FUS, EWSR1 and TAF15 siRNA transfected cells. The cumulative distribution in protein abundance of targeted proteins did not change in statistically significant manner (data not shown). A recent study of FET protein targets also based on PAR-CLIP did not observe FET-target dependent mRNA changes, despite their use of a similar cell system (Hoell et al. 2011). Global changes in protein abundance of 700 to 1100 proteins expressed from FET-targeted mRNAs detected were relatively small (data not shown). However, using the SILAC approach it is only possible to measure changes in protein abundance between siRNA- and mock-transfected cells. On the other hand using pulsed SILAC (pSILAC) it is possible to assess an involvement of FET proteins in protein synthesis (Schwanhausser et al. 2009). To assess the influence of FUS on translation, FUS knockdown under pSILAC conditions was performed. Changes in protein expression from FUS-targeted mRNAs were relatively small (data not shown), indicating that at least FUS on its own is not acting as global translational regulator.

4.4 FET proteins are involved in alternative splicing regulation

Looking further into the detailed context of binding site distribution within introns, an enriched binding for all three FET proteins upstream of 3'splice sites could be observed (Fig.14). This is in line with the assumption that these proteins are involved in pre-mRNA splicing regulation. FUS and EWSR1 were already identified in the same RNA-splicing complex together with polypyrimidine-tract-binding-associated factor (PSF) (Deloulme et al. 1997) and all three proteins were reported to be part of the spliceosomal complex (Rappsilber et al. 2002). Moreover, FUS interacts with different splicing factors like serine/arginine-rich (SR) proteins (Yang et al. 1998). Knockdown of FUS and EWSR1 in HEK293 cells affected splicing of ENAH, THAP6, CSNK1D and PDE8A, respectively (Fig.15). The effect on alternative splicing of target RNAs regulated by FUS and EWSR1 seems to be transcript-specific since exon skipping is both up and down regulated upon FUS and EWSR1 knockdown. Whether FET proteins regulate splicing of target pre-mRNAs directly or just indirectly by recruiting other splicing factors to target mRNAs is not known yet.

Recent reports show similar effects in mouse brain (Lagier-Tourenne et al. 2012; Rogelj et al. 2012). Both studies demonstrated a saw-tooth like binding pattern for FUS in long genes indicating that FUS remains bound to pre-mRNAs until splicing is completed. Furthermore, they did not observe any significant overlap with exons regulated by FUS and TARDBP but both proteins regulate splicing of mRNAs derived from genes involved in neuronal development.

Interestingly, the FUS R495X mutant has less intronic targets due to their cytoplasmic localization and above all less binding sites around splice sites (Fig.7). It is suggested that aberrations of RNA metabolism are important factors in the pathogenesis of ALS and FTLD. It is feasible that aberrantly spliced mRNAs could contribute to the development of ALS. Multiple abnormal mRNAs of the astroglial glutamate transporter protein EAAT2 have been identified in many ALS patients. Proteins translated from these aberrant mRNAs are rapidly degraded and have a dominant negative effect on the wildtype protein leading to a loss of EAAT2 (Lin et al. 1998). EAAT2 deficiency leads to increased extracellular glutamate and neuronal degeneration in motor cortex and spinal cord. Another study also identified aberrantly spliced cell adhesion genes in sporadic ALS patients and some of them could be directly related to TARDBP misregulation (Rabin et al. 2010; Xiao et al. 2011). It may be that FUS performs similar interactions and that disturbance of this function could be linked to ALS.

4.5 FUS and EWSR1 affect mRNA expression of the *SRSF3* locus

It was proposed that the FET family is likely to be involved in regulating transcription by direct interaction with RNA Pol II and TFIID (Bertolotti et al. 1996). In this study, transcriptional regulation by the FET family was investigated using the example of the *SRSF3* region. FUS- and EWSR1-specific binding sites were found downstream of the 3' end of the *SRFS3* transcript (Fig.16). High level of protein binding to this region in general is further supported by protein occupancy profiling (Baltz et al. 2012). In addition RNA-Seq coverage and validation by PCR (data not shown) support existence of such long transcript derived from the *SRFS3* locus which is likely to be a read-through transcript variant of this gene. It is feasible that FUS and EWSR1 regulate transcription of *SRFS3* by interaction with RNA Pol II subunits and their bound mRNA. This is supported by the finding that knockdown of FET proteins leads to an upregulation of the *SRSF3* extended transcript (Fig.16). Whether this is a direct effect of FET protein depletion or just an off-target effect has to be investigated further for example by performing another knockdown using a different set of siRNAs. A recent publication analyzing the genome-wide localization of FUS on the chromatin of HEK293 cells showed an enrichment of FUS at the TSS and an accumulation of RNA Pol II at the TSS upon loss of FUS (Schwartz et al. 2012). Analysis of RNA Pol II occupancy upon triple knockdown around the *SRSF3* gene in this study confirmed these findings (Fig.17). Furthermore, RNA Pol II occupancy decreased at the *SRSF3* 3'UTR and further downstream upon FET protein depletion. This is contradictory to the qPCR data after knockdown where the *SRSF3* extended transcript is upregulated (Fig.16). Since FET proteins are involved in alternative splicing regulation, it is conceivable that FET protein knockdown induces an isoform switch of *SRSF3* producing rather longer than short *SRSF3* transcripts. This needs be verified for example by RT-PCR analysis of *SRSF3* isoform expression or *in vitro* splicing assays. In addition, we tested whether changes in distribution of RNA Pol II were accompanied by changes in Ser2 or Ser5 phosphorylation of the RNA Pol II CTD. In contrast to Schwartz et al., we did not detect any significant changes in Ser2 or Ser5 phosphorylation of RNA Pol II CTD around the *SRSF3* locus upon FET protein knockdown (Fig.17). Schwartz et al. showed that RNA binding nucleates the formation of high-order FUS assemblies which bind the RNA Pol II CTD (Schwartz et al. 2013). Both the LC domain and RGG domain of FUS contribute to the assembly. The authors suggest that the FUS assemblies influence transcription by forming a protein scaffold that recruits RNA Pol II. Another study provided direct evidence that FUS assemblies can stimulate transcription (Kwon et al. 2013). Kwon et al. showed that polymeric fibers formed from the LC domain of FUS, EWSR1 and TAF15 are able bind the RNA Pol II CTD directly.

Phosphorylation of the CTD facilitates the release of the RNA Pol II from the FET LC domain polymer. The authors speculate that the native FET proteins might bind to ncRNAs that remain nascently attached to the DNA bringing these complexes in close proximity to target genes. The locally elevated FET protein concentration might facilitate LC domain polymerization necessary to recruit RNA Pol II.

Furthermore, it would be conceivable that FET proteins link mRNA 3' end processing and transcriptional elongation and termination. The connection between polyadenylation and transcriptional termination was established some time ago discovering that both processes are dependent on the same DNA sequences at 3' ends of genes (Connelly and Manley 1988). Forming the 3' end of a transcript generated by RNA Pol II is a multistep process where the 3' end must be generated by combination of cleavage and polyadenylation. The elongating RNA Pol II complex must be halted and the Pol II enzyme terminated to allow for recycling back to transcriptional initiation (Proudfoot et al. 2002). A strong connection has been made between the phosphorylation of the RNA Pol II CTD and recruitment of polyadenylation factors. For example, the polyadenylation factor Ssu72 for has been shown to have CTD phosphatase activity (Krishnamurthy et al. 2004) raising the possibility that some polyadenylation factors can influence phosphorylation of the CTD to facilitate recycling of the RNA Pol II to a form competent for transcription initiation. Other factors like the FET proteins might also affect directly or indirectly the amount of RNA PoI II at genes thereby regulating the 3'end formation of transcripts.

In conclusion, our data suggest that the FET proteins modulate the total RNA Pol II level at the TSS as well as at the end of the *SRSF3* gene but not the phosphorylation state of the CTD. It would be crucial to investigate RNA Pol II occupancy and CTD phosphorylation on more genomic loci bound by the FET proteins.

4.6 FET proteins are involved in regulation of non-coding RNAs

ncRNAs are involved in many cellular processes and important posttranscriptional regulators of gene expression. This regulation can occur *in cis* or *trans*. Promoter-associated ncRNAs are a group of ncRNAs which can act *in cis*. These transcripts are expressed near transcription start sites or from upstream elements of the promoter. Transcription occurs in both sense and antisense direction with respect to the downstream gene. Most of them are associated with highly expressed genes whereas themselves being only weakly expressed and highly unstable (Preker et al. 2008).

There are several evidences connecting promoter-associated ncRNAs with transcriptional activation and repression. siRNA-directed transcriptional activation of p21 gene expression was demonstrated to be the result of post-transcriptional silencing of a p21-specific antisense transcript (Morris et al. 2008). Expression of ncRNAs derived from the *cyclin D1* (*CCND1*) promoter is induced upon DNA damage (Wang et al. 2008). Binding of these ncRNAs to the 5' regulatory region of the *cyclin D1* gene also recruits FUS to that locus. Upon binding these ncRNAs, FUS changes its allosteric conformation and specifically binds to the CREB-binding protein (CBP) and p300 thus inhibiting their histone acetyltransferase activity and repressing *CCND1* expression. Similarly, EWSR1 and TAF15 were found to bind CBP and p300 and exerted inhibitory effects. We analyzed our FET PAR-CLIP data in combination with RNA-Seq data to find more evidences for this type of transcriptional regulation involving promoter-associated ncRNAs (Fig.18). Multiple putative ncRNAs with FET binding sites in the vicinity of cell cycle promoters were found (Tab.3). This indicates that FET proteins might have a functional role in transcription of cell cycle regulators by being recruited to their loci through gene-specific ncRNAs.

Another group of ncRNAs which act *in trans* are miRNAs which mainly down-regulate expression of target genes. FUS, EWSR1 and TAF15 were already implicated in the regulation miRNA biogenesis, since all three proteins were identified in the same complex with Drosha (Gregory et al. 2004). Our PAR-CLIP results strengthen this assumption as FET proteins have binding sites in several pri-miRNAs (Fig.19Fig. 19).

Knockdown of FUS alone does not seem to have an influence on miRNA expression whereas depletion of all three proteins simultaneously leads to a down regulation of selected miRNAs. This is in agreement with a recent publication where TAF15 depletion lead to decreased pri-miR-17-5p and pri-miR-20a level thereby regulating gene expression of cell cycle regulatory genes (Ballarino et al. 2012). Furthermore, this study showed that expression of Drosha and DGCR8 was unaffected by TAF15 siRNA transfection indicating that this is not an unspecific effect due to a destabilized microprocessor complex. Further work is needed to elucidate whether EWSR1 also participates in the regulation of miRNA biogenesis and if FET proteins maybe work together in the regulation of miRNA processing. Moreover, it would be interesting to look at which step FET proteins regulate miRNA biogenesis. This could be done for instance by quantification of the different precursor miRNA levels using qRT-PCR in FET-depleted cells. Another possibility is that the down regulation of miRNA expression is an unspecific effect since the Drosha complex gets destabilized upon FET depletion.

4.7 Conclusion and outlook

In this study, the role of FET protein members in transcriptional and posttranscriptional gene regulation in comparison with the ALS-related proteins TARDBP and FUS R495X was examined. PAR-CLIP was used to determine the RNA targets of the FET proteins and discovered preferential binding in introns especially near splice sites. Regulation of splicing by FUS and EWSR1 could be verified. Interestingly, the FUS mutant R495X showed reduced binding to splice sites raising the possibility that splicing regulation might be dysregulated in ALS and thus contributing to the pathomechanism. Comparison of FUS and TARDBP RNA targets revealed an overrepresentation of ubiquitin-proteasome-related transcripts which confirms the theory that defects in the protein degradation pathway might be involved in the development of ALS. By using reporter assays targeting different points along the ubiquitin-mediated proteolysis pathway in FUS and TARDBP knockdown cells it will be possible to monitor at which step this pathway might be disordered. Furthermore, by depletion of FUS, EWSR1 and TAF15 it was discovered that FET proteins have stabilizing function of their target RNAs. In addition, FET proteins probably participate in transcriptional regulation on the one hand by interaction with promoter-associated ncRNAs and on the other hand by interaction with RNA Pol II. It would be feasible that FET proteins are connecting transcription and splicing by being recruited to transcribed genes, influencing transcription and remain bound during transcriptional elongation until splicing is completed. Thereby, FET proteins can regulate gene expression of targets at different levels.

Considering the many physiological functions of FUS, EWSR1 and TAF15 in the cell, loss of one or all three proteins due to mutations can alter various cellular processes and have detrimental consequences. Further analysis of FUS and FUS R495X RNA targets will hopefully shed more light in the pathogenesis of ALS and give first starting point in therapy of this severe disease.

References

Aasland R, Gibson TJ, Stewart AF. 1995. The PHD finger: implications for chromatin-mediated transcriptional regulation. *Trends in biochemical sciences* **20**: 56-59.

Anderson P, Kedersha N. 2008. Stress granules: the Tao of RNA triage. *Trends in biochemical sciences* **33**: 141-150.

Andersson MK, Stahlberg A, Arvidsson Y, Olofsson A, Semb H, Stenman G, Nilsson O, Aman P. 2008. The multifunctional FUS, EWS and TAF15 proto-oncoproteins show cell type-specific expression patterns and involvement in cell spreading and stress response. *BMC cell biology* **9**: 37.

Auweter SD, Oberstrass FC, Allain FH. 2006. Sequence-specific binding of single-stranded RNA: is there a code for recognition? *Nucleic acids research* **34**: 4943-4959.

Backe PH, Messias AC, Ravelli RB, Sattler M, Cusack S. 2005. X-ray crystallographic and NMR studies of the third KH domain of hnRNP K in complex with single-stranded nucleic acids. *Structure* **13**: 1055-1067.

Baechtold H, Kuroda M, Sok J, Ron D, Lopez BS, Akhmedov AT. 1999. Human 75-kDa DNA-pairing protein is identical to the pro-oncoprotein TLS/FUS and is able to promote D-loop formation. *The Journal of biological chemistry* **274**: 34337-34342.

Ballarino M, Jobert L, Dembele D, de la Grange P, Auboeuf D, Tora L. 2012. TAF15 is important for cellular proliferation and regulates the expression of a subset of cell cycle genes through miRNAs. *Oncogene*.

Baltz AG, Munschauer M, Schwanhausser B, Vasile A, Murakawa Y, Schueler M, Youngs N, Penfold-Brown D, Drew K, Milek M et al. 2012. The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Molecular cell* **46**: 674-690.

Bertolotti A, Bell B, Tora L. 1999. The N-terminal domain of human TAFII68 displays transactivation and oncogenic properties. *Oncogene* **18**: 8000-8010.

Bertolotti A, Lutz Y, Heard DJ, Chambon P, Tora L. 1996. hTAF(II)68, a novel RNA/ssDNA-binding protein with homology to the pro-oncoproteins TLS/FUS and EWS is associated with both TFIID and RNA polymerase II. *The EMBO journal* **15**: 5022-5031.

Bertolotti A, Melot T, Acker J, Vigneron M, Delattre O, Tora L. 1998. EWS, but not EWS-FLI-1, is associated with both TFIID and RNA polymerase II: interactions between two members of the TET family, EWS and hTAFII68, and subunits of TFIID and RNA polymerase II complexes. *Molecular and cellular biology* **18**: 1489-1497.

Bertrand P, Akhmedov AT, Delacote F, Durrbach A, Lopez BS. 1999. Human POMp75 is identified as the pro-oncoprotein TLS/FUS: both POMp75 and POMp100 DNA homologous pairing activities are associated to cell proliferation. *Oncogene* **18**: 4515-4521.

Bosco DA, Lemay N, Ko HK, Zhou H, Burke C, Kwiatkowski TJ, Jr., Sapp P, McKenna-Yasek D, Brown RH, Jr., Hayward LJ. 2010. Mutant FUS proteins that cause amyotrophic lateral sclerosis incorporate into stress granules. *Human molecular genetics* **19**: 4160-4175.

Brettschneider J, Van Deerlin VM, Robinson JL, Kwong L, Lee EB, Ali YO, Safren N, Monteiro MJ, Toledo JB, Elman L et al. 2012. Pattern of ubiquilin pathology in ALS and FTLD indicates presence of C9ORF72 hexanucleotide expansion. *Acta neuropathologica* **123**: 825-839.

Buratti E, Baralle FE. 2004. Influence of RNA secondary structure on the pre-mRNA splicing process. *Molecular and cellular biology* **24**: 10505-10514.

Buratti E, Brindisi A, Giombi M, Tisminetzky S, Ayala YM, Baralle FE. 2005. TDP-43 binds heterogeneous nuclear ribonucleoprotein A/B through its C-terminal tail: an important region for the inhibition of cystic fibrosis transmembrane conductance regulator exon 9 splicing. *The Journal of biological chemistry* **280**: 37572-37584.

Burd CG, Dreyfuss G. 1994. Conserved structures and diversity of functions of RNA-binding proteins. *Science* **265**: 615-621.

Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL. 2011. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev* **25**: 1915-1927.

Calvio C, Neubauer G, Mann M, Lamond AI. 1995. Identification of hnRNP P2 as TLS/FUS using electrospray mass spectrometry. *Rna* **1**: 724-733.

Castello A, Fischer B, Eichelbaum K, Horos R, Beckmann BM, Strein C, Davey NE, Humphreys DT, Preiss T, Steinmetz LM et al. 2012. Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* **149**: 1393-1406.

Chang KY, Ramos A. 2005. The double-stranded RNA-binding motif, a versatile macromolecular docking platform. *The FEBS journal* **272**: 2109-2117.

Chansky HA, Hu M, Hickstein DD, Yang L. 2001. Oncogenic TLS/ERG and EWS/Fli-1 fusion proteins inhibit RNA splicing mediated by YB-1 protein. *Cancer research* **61**: 3586-3590.

Couthouis J, Hart MP, Erion R, King OD, Diaz Z, Nakaya T, Ibrahim F, Kim HJ, Mojsilovic-Petrovic J, Panossian S et al. 2012. Evaluating the role of the FUS/TLS-related gene EWSR1 in amyotrophic lateral sclerosis. *Human molecular genetics* **21**: 2899-2911.

Couthouis J, Hart MP, Shorter J, DeJesus-Hernandez M, Erion R, Oristano R, Liu AX, Ramos D, Jethava N, Hosangadi D et al. 2011. A yeast functional screen predicts new candidate ALS disease genes. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 20881-20890.

Crozat A, Aman P, Mandahl N, Ron D. 1993. Fusion of CHOP to a novel RNA-binding protein in human myxoid liposarcoma. *Nature* **363**: 640-644.

Daigle JG, Lanson NA, Jr., Smith RB, Casci I, Maltare A, Monaghan J, Nichols CD, Kryndushkin D, Shewmaker F, Pandey UB. 2013. RNA-binding ability of FUS regulates

neurodegeneration, cytoplasmic mislocalization and incorporation into stress granules associated with FUS carrying ALS-linked mutations. *Human molecular genetics* **22**: 1193-1205.

de Hoog CL, Foster LJ, Mann M. 2004. RNA and RNA binding proteins participate in early stages of cell spreading through spreading initiation centers. *Cell* **117**: 649-662.

de Lima Morais DA, Fang H, Rackham OJ, Wilson D, Pethica R, Chothia C, Gough J. 2011. SUPERFAMILY 1.75 including a domain-centric gene ontology method. *Nucleic acids research* **39**: D427-434.

Del Bo R, Tiloca C, Pensato V, Corrado L, Ratti A, Ticozzi N, Corti S, Castellotti B, Mazzini L, Soraru G et al. 2011. Novel optineurin mutations in patients with familial and sporadic amyotrophic lateral sclerosis. *Journal of neurology, neurosurgery, and psychiatry* **82**: 1239-1243.

Delattre O, Zucman J, Plougastel B, Desmaze C, Melot T, Peter M, Kovar H, Joubert I, de Jong P, Rouleau G et al. 1992. Gene fusion with an ETS DNA-binding domain caused by chromosome translocation in human tumours. *Nature* **359**: 162-165.

Deloulme JC, Prichard L, Delattre O, Storm DR. 1997. The prooncoprotein EWS binds calmodulin and is phosphorylated by protein kinase C through an IQ domain. *The Journal of biological chemistry* **272**: 27369-27377.

Deng HX, Chen W, Hong ST, Boycott KM, Gorrie GH, Siddique N, Yang Y, Fecto F, Shi Y, Zhai H et al. 2011. Mutations in UBQLN2 cause dominant X-linked juvenile and adult-onset ALS and ALS/dementia. *Nature* **477**: 211-215.

Dormann D, Rodde R, Edbauer D, Bentmann E, Fischer I, Hruscha A, Than ME, Mackenzie IR, Capell A, Schmid B et al. 2010. ALS-associated fused in sarcoma (FUS) mutations disrupt Transportin-mediated nuclear import. *The EMBO journal* **29**: 2841-2857.

Dreyfuss G, Kim VN, Kataoka N. 2002. Messenger-RNA-binding proteins and the messages they carry. *Nature reviews Molecular cell biology* **3**: 195-205.

Fecto F, Yan J, Vemula SP, Liu E, Yang Y, Chen W, Zheng JG, Shi Y, Siddique N, Arrat H et al. 2011. SQSTM1 mutations in familial and sporadic amyotrophic lateral sclerosis. *Archives of neurology* **68**: 1440-1446.

Fitzgerald KD, Semler BL. 2011. Re-localization of cellular protein SRp20 during poliovirus infection: bridging a viral IRES to the host cell translation apparatus. *PLoS pathogens* **7**: e1002127.

Fujii R, Okabe S, Urushido T, Inoue K, Yoshimura A, Tachibana T, Nishikawa T, Hicks GG, Takumi T. 2005. The RNA binding protein TLS is translocated to dendritic spines by mGluR5 activation and regulates spine morphology. *Current biology : CB* **15**: 587-593.

Fujii R, Takumi T. 2005. TLS facilitates transport of mRNA encoding an actin-stabilizing protein to dendritic spines. *Journal of cell science* **118**: 5755-5765.

Gamsjaeger R, Liew CK, Loughlin FE, Crossley M, Mackay JP. 2007. Sticky fingers: zinc-fingers as protein-recognition motifs. *Trends in biochemical sciences* **32**: 63-70.

Gregory RI, Yan KP, Amuthan G, Chendrimada T, Doratotaj B, Cooch N, Shiekhattar R. 2004. The Microprocessor complex mediates the genesis of microRNAs. *Nature* **432**: 235-240.

Grishin NV. 2001. KH domain: one motif, two folds. *Nucleic acids research* **29**: 638-643.

Guipaud O, Guillonneau F, Labas V, Praseuth D, Rossier J, Lopez B, Bertrand P. 2006. An in vitro enzymatic assay coupled to proteomics analysis reveals a new DNA processing activity for Ewing sarcoma and TAF(II)68 proteins. *Proteomics* **6**: 5962-5972.

Hackl W, Luhrmann R. 1996. Molecular cloning and subcellular localisation of the snRNP-associated protein 69KD, a structural homologue of the proto-oncoproteins TLS and EWS with RNA and DNA-binding properties. *Journal of molecular biology* **264**: 843-851.

Hafner M, Landgraf P, Ludwig J, Rice A, Ojo T, Lin C, Holoch D, Lim C, Tuschl T. 2008. Identification of microRNAs and other small regulatory RNAs using cDNA library sequencing. *Methods* **44**: 3-12.

Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M, Jr., Jungkamp AC, Munschauer M et al. 2010. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**: 129-141.

Hall TM. 2005. Multiple modes of RNA recognition by zinc finger proteins. *Current opinion in structural biology* **15**: 367-373.

Hallier M, Lerga A, Barnache S, Tavitian A, Moreau-Gachelin F. 1998. The transcription factor Spi-1/PU.1 interacts with the potential splicing factor TLS. *The Journal of biological chemistry* **273**: 4838-4842.

Han TW, Kato M, Xie S, Wu LC, Mirzaei H, Pei J, Chen M, Xie Y, Allen J, Xiao G et al. 2012. Cell-free formation of RNA granules: bound RNAs identify features and components of cellular assemblies. *Cell* **149**: 768-779.

Hicks GG, Singh N, Nashabi A, Mai S, Bozek G, Klewes L, Arapovic D, White EK, Koury MJ, Oltz EM et al. 2000. Fus deficiency in mice results in defective B-lymphocyte development and activation, high levels of chromosomal instability and perinatal death. *Nature genetics* **24**: 175-179.

Hoell JI, Larsson E, Runge S, Nusbaum JD, Duggimpudi S, Farazi TA, Hafner M, Borkhardt A, Sander C, Tuschl T. 2011. RNA targets of wild-type and mutant FET family proteins. *Nature structural & molecular biology* **18**: 1428-1431.

Hofacker IL. 2004. RNA secondary structure analysis using the Vienna RNA package. *Current protocols in bioinformatics / editoral board, Andreas D Baxevanis [et al]* **Chapter 12**: Unit 12 12.

Hoffmann A, Roeder RG. 1996. Cloning and characterization of human TAF20/15. Multiple interactions suggest a central role in TFIID complex formation. *The Journal of biological chemistry* **271**: 18194-18202.

Huang da W, Sherman BT, Lempicki RA. 2009a. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic acids research* **37**: 1-13.

-. 2009b. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* **4**: 44-57.

Huang Y, Steitz JA. 2001. Splicing factors SRp20 and 9G8 promote the nucleocytoplasmic export of mRNA. *Molecular cell* **7**: 899-905.

Ibrahim F, Maragkakis M, Alexiou P, Maronski MA, Dichter MA, Mourelatos Z. 2013. Identification of in vivo, conserved, TAF15 RNA binding sites reveals the impact of TAF15 on the neuronal transcriptome. *Cell reports* **3**: 301-308.

Ishigaki S, Masuda A, Fujioka Y, Iguchi Y, Katsuno M, Shibata A, Urano F, Sobue G, Ohno K. 2012. Position-dependent FUS-RNA interactions regulate alternative splicing events and transcriptions. *Scientific reports* **2**: 529.

Ito D, Seki M, Tsunoda Y, Uchiyama H, Suzuki N. 2011. Nuclear transport impairment of amyotrophic lateral sclerosis-linked mutations in FUS/TLS. *Annals of neurology* **69**: 152-162.

Jia R, Li C, McCoy JP, Deng CX, Zheng ZM. 2010. SRp20 is a proto-oncogene critical for cell proliferation and tumor induction and maintenance. *International journal of biological sciences* **6**: 806-826.

Jobert L, Pinzon N, Van Herreweghe E, Jady BE, Guialis A, Kiss T, Tora L. 2009. Human U1 snRNA forms a new chromatin-associated snRNP with TAF15. *EMBO reports* **10**: 494-500.

Johnson JO, Mandrioli J, Benatar M, Abramzon Y, Van Deerlin VM, Trojanowski JQ, Gibbs JR, Brunetti M, Gronka S, Wuu J et al. 2010. Exome sequencing reveals VCP mutations as a cause of familial ALS. *Neuron* **68**: 857-864.

Kabashi E, Valdmanis PN, Dion P, Spiegelman D, McConkey BJ, Vande Velde C, Bouchard JP, Lacomblez L, Pochigaeva K, Salachas F et al. 2008. TARDBP mutations in individuals with sporadic and familial amyotrophic lateral sclerosis. *Nature genetics* **40**: 572-574.

Kahvejian A, Svitkin YV, Sukarieh R, M'Boutchou MN, Sonenberg N. 2005. Mammalian poly(A)-binding protein is a eukaryotic translation initiation factor, which acts via multiple mechanisms. *Genes & development* **19**: 104-113.

Kato M, Han TW, Xie S, Shi K, Du X, Wu LC, Mirzaei H, Goldsmith EJ, Longgood J, Pei J et al. 2012. Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149**: 753-767.

Keene JD. 2007. RNA regulons: coordination of post-transcriptional events. *Nature reviews Genetics* **8**: 533-543.

Kiernan MC, Vucic S, Cheah BC, Turner MR, Eisen A, Hardiman O, Burrell JR, Zoing MC. 2011. Amyotrophic lateral sclerosis. *Lancet* **377**: 942-955.

Kim SH, Shanware NP, Bowler MJ, Tibbetts RS. 2010. Amyotrophic lateral sclerosis-associated proteins TDP-43 and FUS/TLS function in a common biochemical complex to co-regulate HDAC6 mRNA. *The Journal of biological chemistry* **285**: 34097-34105.

Kino Y, Washizu C, Aquilanti E, Okuno M, Kurosawa M, Yamada M, Doi H, Nukina N. 2011. Intracellular localization and splicing regulation of FUS/TLS are variably affected by amyotrophic lateral sclerosis-linked mutations. *Nucleic acids research* **39**: 2781-2798.

Komarnitsky P, Cho EJ, Buratowski S. 2000. Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes & development* **14**: 2452-2460.

Krishnamurthy S, He X, Reyes-Reyes M, Moore C, Hampsey M. 2004. Ssu72 Is an RNA polymerase II CTD phosphatase. *Molecular cell* **14**: 387-394.

Kuroda M, Sok J, Webb L, Baechtold H, Urano F, Yin Y, Chung P, de Rooij DG, Akhmedov A, Ashley T et al. 2000. Male sterility and enhanced radiation sensitivity in TLS(-/-) mice. *The EMBO journal* **19**: 453-462.

Kwiatkowski TJ, Jr., Bosco DA, Leclerc AL, Tamrazian E, Vanderburg CR, Russ C, Davis A, Gilchrist J, Kasarskis EJ, Munsat T et al. 2009. Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. *Science* **323**: 1205-1208.

Kwon I, Kato M, Xiang S, Wu L, Theodoropoulos P, Mirzaei H, Han T, Xie S, Corden JL, McKnight SL. 2013. Phosphorylation-regulated binding of RNA polymerase II to fibrous polymers of low-complexity domains. *Cell* **155**: 1049-1060.

Lagier-Tourenne C, Polymenidou M, Hutt KR, Vu AQ, Baughn M, Huelga SC, Clutario KM, Ling SC, Liang TY, Mazur C et al. 2012. Divergent roles of ALS-linked proteins FUS/TLS and TDP-43 intersect in processing long pre-mRNAs. *Nature neuroscience* **15**: 1488-1497.

Lebedeva S, Jens M, Theil K, Schwanhausser B, Selbach M, Landthaler M, Rajewsky N. 2011. Transcriptome-wide analysis of regulatory interactions of the RNA-binding protein HuR. *Mol Cell* **43**: 340-352.

Lee J, Rhee BK, Bae GY, Han YM, Kim J. 2005. Stimulation of Oct-4 activity by Ewing's sarcoma protein. *Stem cells* **23**: 738-751.

Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Radmark O, Kim S et al. 2003. The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**: 415-419.

Leemann-Zakaryan RP, Pahlich S, Grossenbacher D, Gehring H. 2011. Tyrosine Phosphorylation in the C-Terminal Nuclear Localization and Retention Signal (C-NLS) of the EWS Protein. *Sarcoma* **2011**: 218483.

Lerga A, Hallier M, Delva L, Orvain C, Gallais I, Marie J, Moreau-Gachelin F. 2001. Identification of an RNA binding specificity for the potential splicing factor TLS. *The Journal of biological chemistry* **276**: 6807-6816.

Lewis HA, Chen H, Edo C, Buckanovich RJ, Yang YY, Musunuru K, Zhong R, Darnell RB, Burley SK. 1999. Crystal structures of Nova-1 and Nova-2 K-homology RNA-binding domains. *Structure* **7**: 191-203.

Li H, Watford W, Li C, Parmelee A, Bryant MA, Deng C, O'Shea J, Lee SB. 2007. Ewing sarcoma gene EWS is essential for meiosis and B lymphocyte development. *The Journal of clinical investigation* **117**: 1314-1323.

Lin CL, Bristol LA, Jin L, Dykes-Hoberg M, Crawford T, Clawson L, Rothstein JD. 1998. Aberrant RNA processing in a neurodegenerative disease: the cause for absent EAAT2, a glutamate transporter, in amyotrophic lateral sclerosis. *Neuron* **20**: 589-602.

Ling SC, Albuquerque CP, Han JS, Lagier-Tourenne C, Tokunaga S, Zhou H, Cleveland DW. 2010. ALS-associated mutations in TDP-43 increase its stability and promote TDP-43 complexes with FUS/TLS. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 13318-13323.

Loughlin FE, Mansfield RE, Vaz PM, McGrath AP, Setiyaputra S, Gamsjaeger R, Chen ES, Morris BJ, Guss JM, Mackay JP. 2009. The zinc fingers of the SR-like protein ZRANB2 are single-stranded RNA-binding domains that recognize 5' splice site-like sequences. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 5581-5586.

Lu Z, Xu S, Joazeiro C, Cobb MH, Hunter T. 2002. The PHD domain of MEKK1 acts as an E3 ubiquitin ligase and mediates ubiquitination and degradation of ERK1/2. *Molecular cell* **9**: 945-956.

Lunde BM, Moore C, Varani G. 2007. RNA-binding proteins: modular design for efficient function. *Nature reviews Molecular cell biology* **8**: 479-490.

Marko M, Vlassis A, Guialis A, Leichter M. 2012. Domains involved in TAF15 subcellular localisation: dependence on cell type and ongoing transcription. *Gene* **506**: 331-338.

Martini A, La Starza R, Janssen H, Bilhou-Nabera C, Corveleyn A, Somers R, Aventin A, Foa R, Hagemeijer A, Mecucci C et al. 2002. Recurrent rearrangement of the Ewing's sarcoma gene, EWSR1, or its homologue, TAF15, with the transcription factor CIZ/NMP4 in acute leukemia. *Cancer research* **62**: 5408-5412.

Maruyama H, Morino H, Ito H, Izumi Y, Kato H, Watanabe Y, Kinoshita Y, Kamada M, Nodera H, Suzuki H et al. 2010. Mutations of optineurin in amyotrophic lateral sclerosis. *Nature* **465**: 223-226.

Matthews JM, Sunde M. 2002. Zinc fingers--folds for many occasions. *IUBMB life* **54**: 351-355.

Mazza C, Segref A, Mattaj IW, Cusack S. 2002. Large-scale induced fit recognition of an m(7)GpppG cap analogue by the human nuclear cap-binding complex. *The EMBO journal* **21**: 5548-5557.

Meissner M, Lopato S, Gotzmann J, Sauermann G, Barta A. 2003. Proto-oncoprotein TLS/FUS is associated to the nuclear matrix and complexed with splicing factors PTB, SRm160, and SR proteins. *Experimental cell research* **283**: 184-195.

Meyer H, Bug M, Bremer S. 2012. Emerging functions of the VCP/p97 AAA-ATPase in the ubiquitin system. *Nature cell biology* **14**: 117-123.

Michelitsch MD, Weissman JS. 2000. A census of glutamine/asparagine-rich regions: implications for their conserved function and the prediction of novel prions. *Proceedings of the National Academy of Sciences of the United States of America* **97**: 11910-11915.

Miller J, McLachlan AD, Klug A. 1985. Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. *The EMBO journal* **4**: 1609-1614.

Montes M, Becerra S, Sanchez-Alvarez M, Sune C. 2012. Functional coupling of transcription and splicing. *Gene* **501**: 104-117.

Morlando M, Dini Modigliani S, Torrelli G, Rosa A, Di Carlo V, Caffarelli E, Bozzoni I. 2012. FUS stimulates microRNA biogenesis by facilitating co-transcriptional Drosha recruitment. *The EMBO journal* **31**: 4502-4510.

Morris DP, Michelotti GA, Schwinn DA. 2005. Evidence that phosphorylation of the RNA polymerase II carboxyl-terminal repeats is similar in yeast and humans. *The Journal of biological chemistry* **280**: 31368-31377.

Morris KV, Santoso S, Turner AM, Pastori C, Hawkins PG. 2008. Bidirectional transcription directs both transcriptional gene activation and suppression in human cells. *PLoS genetics* **4**: e1000258.

Moscat J, Diaz-Meco MT. 2012. p62: a versatile multitasker takes on cancer. *Trends in biochemical sciences* **37**: 230-236.

Neumann M, Bentmann E, Dormann D, Jawaid A, DeJesus-Hernandez M, Ansorge O, Roeber S, Kretzschmar HA, Munoz DG, Kusaka H et al. 2011. FET proteins TAF15 and EWS are selective markers that distinguish FTLD with FUS pathology from amyotrophic lateral sclerosis with FUS mutations. *Brain : a journal of neurology* **134**: 2595-2609.

Neumann M, Rademakers R, Roeber S, Baker M, Kretzschmar HA, Mackenzie IR. 2009. A new subtype of frontotemporal lobar degeneration with FUS pathology. *Brain* **132**: 2922-2931.

Neumann M, Sampathu DM, Kwong LK, Truax AC, Micsenyi MC, Chou TT, Bruce J, Schuck T, Grossman M, Clark CM et al. 2006. Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science* **314**: 130-133.

Nguyen CD, Mansfield RE, Leung W, Vaz PM, Loughlin FE, Grant RP, Mackay JP. 2011. Characterization of a family of RanBP2-type zinc fingers that can recognize single-stranded RNA. *Journal of molecular biology* **407**: 273-283.

Ohno T, Ouchida M, Lee L, Gatalica Z, Rao VN, Reddy ES. 1994. The EWS gene, involved in Ewing family of tumors, malignant melanoma of soft parts and desmoplastic small round cell tumors, codes for an RNA binding protein with novel regulatory domains. *Oncogene* **9**: 3087-3097.

Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M. 2002. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Molecular & cellular proteomics : MCP* **1**: 376-386.

Oubridge C, Ito N, Evans PR, Teo CH, Nagai K. 1994. Crystal structure at 1.92 A resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature* **372**: 432-438.

Paronetto MP, Minana B, Valcarcel J. 2011. The Ewing sarcoma protein regulates DNA damage-induced alternative splicing. *Molecular cell* **43**: 353-368.

Polymenidou M, Lagier-Tourenne C, Hutt KR, Huelga SC, Moran J, Liang TY, Ling SC, Sun E, Wancewicz E, Mazur C et al. 2011. Long pre-mRNA depletion and RNA missplicing contribute to neuronal vulnerability from loss of TDP-43. *Nat Neurosci* **14**: 459-468.

Preker P, Nielsen J, Kammler S, Lykke-Andersen S, Christensen MS, Mapendano CK, Schierup MH, Jensen TH. 2008. RNA exosome depletion reveals transcription upstream of active human promoters. *Science* **322**: 1851-1854.

Price SR, Evans PR, Nagai K. 1998. Crystal structure of the spliceosomal U2B"-U2A' protein complex bound to a fragment of U2 small nuclear RNA. *Nature* **394**: 645-650.

Proudfoot NJ, Furger A, Dye MJ. 2002. Integrating mRNA processing with transcription. *Cell* **108**: 501-512.

Pruitt KD, Tatusova T, Maglott DR. 2005. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic acids research* **33**: D501-504.

Rabbitts TH, Forster A, Larson R, Nathan P. 1993. Fusion of the dominant negative transcription regulator CHOP with a novel gene FUS by translocation t(12;16) in malignant liposarcoma. *Nature genetics* **4**: 175-180.

Rabin SJ, Kim JM, Baughn M, Libby RT, Kim YJ, Fan Y, Libby RT, La Spada A, Stone B, Ravits J. 2010. Sporadic ALS has compartment-specific aberrant exon splicing and altered cell-matrix adhesion biology. *Human molecular genetics* **19**: 313-328.

Rappsilber J, Ryder U, Lamond AI, Mann M. 2002. Large-scale proteomic analysis of the human spliceosome. *Genome research* **12**: 1231-1245.

Ray D, Kazan H, Cook KB, Weirauch MT, Najafabadi HS, Li X, Gueroussov S, Albu M, Zheng H, Yang A et al. 2013. A compendium of RNA-binding motifs for decoding gene regulation. *Nature* **499**: 172-177.

Reed R. 1989. The organization of 3' splice-site sequences in mammalian introns. *Genes & development* **3**: 2113-2123.

Rogelj B, Easton LE, Bogu GK, Stanton LW, Rot G, Curk T, Zupan B, Sugimoto Y, Modic M, Haberman N et al. 2012. Widespread binding of FUS along nascent RNA regulates alternative splicing in the brain. *Scientific reports* **2**: 603.

Rubino E, Rainero I, Chio A, Rogaeva E, Galimberti D, Fenoglio P, Grinberg Y, Isaia G, Calvo A, Gentile S et al. 2012. SQSTM1 mutations in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Neurology* **79**: 1556-1562.

Sasagawa K, Matsudo Y, Kang M, Fujimura L, Iitsuka Y, Okada S, Ochiai T, Tokuhisa T, Hatano M. 2002. Identification of Nd1, a novel murine kelch family protein, involved in stabilization of actin filaments. *The Journal of biological chemistry* **277**: 44140-44146.

Schwanhausser B, Gossen M, Dittmar G, Selbach M. 2009. Global analysis of cellular protein translation by pulsed SILAC. *Proteomics* **9**: 205-209.

Schwartz JC, Ebmeier CC, Podell ER, Heimiller J, Taatjes DJ, Cech TR. 2012. FUS binds the CTD of RNA polymerase II and regulates its phosphorylation at Ser2. *Genes & development* **26**: 2690-2695.

Schwartz JC, Wang X, Podell ER, Cech TR. 2013. RNA Seeds Higher-Order Assembly of FUS Protein. *Cell reports* **5**: 918-925.

Shelkovnikova TA, Peters OM, Deykin AV, Connor-Robson N, Robinson H, Ustyugov AA, Bachurin SO, Ermolkevich TG, Goldman IL, Sadchikova ER et al. 2013. Fused in sarcoma (FUS) protein lacking nuclear localization signal (NLS) and major RNA binding motifs triggers proteinopathy and severe motor phenotype in transgenic mice. *The Journal of biological chemistry* **288**: 25266-25274.

Shiohama A, Sasaki T, Noda S, Minoshima S, Shimizu N. 2007. Nucleolar localization of DGCR8 and identification of eleven DGCR8-associated proteins. *Experimental cell research* **313**: 4196-4207.

Sohn EJ, Park J, Kang SI, Wu YP. 2012. Accumulation of pre-let-7g and downregulation of mature let-7g with the depletion of EWS. *Biochemical and biophysical research communications* **426**: 89-93.

Spahn L, Petermann R, Siligan C, Schmid JA, Aryee DN, Kovar H. 2002. Interaction of the EWS NH2 terminus with BARD1 links the Ewing's sarcoma gene to a common tumor suppressor pathway. *Cancer research* **62**: 4583-4587.

Spitzer JI, Ugras S, Runge S, Decarolis P, Antonescu C, Tuschl T, Singer S. 2011. mRNA and protein levels of FUS, EWSR1, and TAF15 are upregulated in liposarcoma. *Genes, chromosomes & cancer* **50**: 338-347.

Sreedharan J, Blair IP, Tripathi VB, Hu X, Vance C, Rogelj B, Ackerley S, Durnall JC, Williams KL, Buratti E et al. 2008. TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis. *Science* **319**: 1668-1672.

Stefl R, Oberstrass FC, Hood JL, Jourdan M, Zimmermann M, Skrisovska L, Maris C, Peng L, Hofr C, Emeson RB et al. 2010. The solution structure of the ADAR2 dsRBM-RNA complex reveals a sequence-specific readout of the minor groove. *Cell* **143**: 225-237.

Stefl R, Skrisovska L, Allain FH. 2005. RNA sequence- and shape-dependent recognition by proteins in the ribonucleoprotein particle. *EMBO reports* **6**: 33-38.

Takahama K, Takada A, Tada S, Shimizu M, Sayama K, Kurokawa R, Oyoshi T. 2013. Regulation of telomere length by G-quadruplex telomere DNA- and TERRA-binding protein TLS/FUS. *Chemistry & biology* **20**: 341-350.

Tan AY, Manley JL. 2009. The TET family of proteins: functions and roles in disease. *Journal of molecular cell biology* **1**: 82-92.

-. 2010. TLS inhibits RNA polymerase III transcription. *Molecular and cellular biology* **30**: 186-196.

Teyssou E, Takeda T, Lebon V, Boillee S, Doukoure B, Bataillon G, Sazdovitch V, Cazeneuve C, Meininger V, LeGuern E et al. 2013. Mutations in SQSTM1 encoding p62 in amyotrophic lateral sclerosis: genetics and neuropathology. *Acta neuropathologica* **125**: 511-522.

Thomas GR, Latchman DS. 2002. The pro-oncoprotein EWS (Ewing's Sarcoma protein) interacts with the Brn-3a POU transcription factor and inhibits its ability to activate transcription. *Cancer biology & therapy* **1**: 428-432.

Ticozzi N, Vance C, Leclerc AL, Keagle P, Glass JD, McKenna-Yasek D, Sapp PC, Silani V, Bosco DA, Shaw CE et al. 2011. Mutational analysis reveals the FUS homolog TAF15 as a candidate gene for familial amyotrophic lateral sclerosis. *American journal of medical genetics Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics* **156B**: 285-290.

Tollervey JR, Curk T, Rogelj B, Briese M, Cereda M, Kayikci M, Konig J, Hortobagyi T, Nishimura AL, Zupunski V et al. 2011. Characterizing the RNA targets and position-dependent splicing regulation by TDP-43. *Nat Neurosci* **14**: 452-458.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**: 1105-1111.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature biotechnology* **28**: 511-515.

Vance C, Rogelj B, Hortobagyi T, De Vos KJ, Nishimura AL, Sreedharan J, Hu X, Smith B, Ruddy D, Wright P et al. 2009. Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. *Science* **323**: 1208-1211.

Venkitaraman AR. 2001. Functions of BRCA1 and BRCA2 in the biological response to DNA damage. *Journal of cell science* **114**: 3591-3598.

Wang X, Arai S, Song X, Reichart D, Du K, Pascual G, Tempst P, Rosenfeld MG, Glass CK, Kurokawa R. 2008. Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. *Nature* **454**: 126-130.

Workman JL, Roeder RG. 1987. Binding of transcription factor TFIID to the major late promoter during in vitro nucleosome assembly potentiates subsequent initiation by RNA polymerase II. *Cell* **51**: 613-622.

Wu S, Green MR. 1997. Identification of a human protein that recognizes the 3' splice site during the second step of pre-mRNA splicing. *The EMBO journal* **16**: 4421-4432.

Xiao S, Sanelli T, Dib S, Sheps D, Findlater J, Bilbao J, Keith J, Zinman L, Rogaeva E, Robertson J. 2011. RNA targets of TDP-43 identified by UV-CLIP are deregulated in ALS. *Molecular and cellular neurosciences* **47**: 167-180.

Yang L, Chansky HA, Hickstein DD. 2000. EWS.Fli-1 fusion protein interacts with hyperphosphorylated RNA polymerase II and interferes with serine-arginine protein-mediated RNA splicing. *The Journal of biological chemistry* **275**: 37612-37618.

Yang L, Embree LJ, Tsai S, Hickstein DD. 1998. Oncoprotein TLS interacts with serine-arginine proteins involved in RNA splicing. *The Journal of biological chemistry* **273**: 27761-27764.

Yeom KH, Lee Y, Han J, Suh MR, Kim VN. 2006. Characterization of DGCR8/Pasha, the essential cofactor for Drosha in primary miRNA processing. *Nucleic acids research* **34**: 4622-4629.

Zakaryan RP, Gehring H. 2006. Identification and characterization of the nuclear localization/retention signal in the EWS proto-oncoprotein. *Journal of molecular biology* **363**: 27-38.

Zinszner H, Albalat R, Ron D. 1994. A novel effector domain from the RNA-binding protein TLS or EWS is required for oncogenic transformation by CHOP. *Genes & development* **8**: 2513-2526.

Zinszner H, Sok J, Immanuel D, Yin Y, Ron D. 1997. TLS (FUS) binds RNA in vivo and engages in nucleo-cytoplasmic shuttling. *Journal of cell science* **110 (Pt 15)**: 1741-1750.

Zucman J, Melot T, Desmaze C, Ghysdael J, Plougastel B, Peter M, Zucker JM, Triche TJ, Sheer D, Turc-Carel C et al. 1993. Combinatorial generation of variable fusion proteins in the Ewing family of tumours. *The EMBO journal* **12**: 4481-4487.

List of abbreviations

4SU	4-thiouridine
6SG	6-thioguanosine
AE	amplification efficiency
ALS	amyotrophic lateral sclerosis
ATP	adenosine triphosphate
BARD1	BRCA1-associated ring finger domain protein 1
bp	base pairs
BRCA1	breast cancer 1
BWA	Burrows-Wheeler Alignment tool
C9orf72	chromosome 9 open reading frame 72
CBP	CREB-binding protein
CCND1	cyclin D1
CCR	crosslink centered region
CDF	cumulative distribution function
CDK1	cyclin-dependent kinase 1
CDKN1A	cyclin-dependent kinase inhibitor 1A
cDNA	complementary DNA
CDS	coding sequence
ChIP	chromatin immunoprecipitation
CHOP	CCAAT/enhancer-binding homologous protein
CSNK1D	casein kinase 1 delta
C_t	cycle treshold
CTD	carboxy terminal domain

C-terminus	carboxy-terminus
DGCR8	DiGeorge syndrome critical region 8
DMEM	Dulbecco's modified Eagle's medium
DNA	deoxyribonucleic acid
ds	double-stranded
dsRBM	double-stranded RNA-binding motif
EAA2	excitatory amino acid transporter 2
EDTA	Ethylenediaminetetraacetic acid
EMSA	electrophoretic mobility shift assay
ENAH	enabled homolog
ERG	v-ets avian erythroblastosis virus E26 oncogene homolog
EWSR1	EWS RNA-binding protein 1
FBS	fetal bovine serum
FDR	False discovery rate
FET family	FUS, EWSR1, TAF15 family
FPKM	fragments per kilobase of transcript per million mapped reads
FTLD	frontemporal lobar degeneration
FUS	fused in sarcoma
GAPDH	glyceraldehyde-3-phosphate dehydrogenase
GO	gene ontology
HA	hemagglutinin
HDAC6	histone deacetylase 6
HEK293 cells	Human Embryonic Kidney 293 cells
HITS-CLIP	high-throughput sequencing of RNA isolated by crosslinking immunoprecipitation

hnRNP	heterogenous nuclear ribonucleoprotein
HRP	horseradish peroxidase
iCLIP	individual-nucleotide resolution crosslinking and immunoprecipitation
IP	immunoprecipitation
IRES	internal ribosome entry side
KD	knockdown
KH domain	K-homology domain
LFC	Log 2 fold changes
LC sequences	Low complexity sequences
LC-MS/MS	liquid chromatography tandem mass spectrometry
LIG1	DNA ligase 1
lincRNA	long intergenic non-coding RNA
miRNA	microRNA
mRNA	messenger RNA
ncRNA	non-coding RNA
NLS	nuclear localization signal
nt	nucleotides
N-terminus	amino-terminus
NUCKS1	nuclear casein kinase and cyclin-dependent kinase substrate 1
OCT4	octamer-binding protein 4
OPTN	optineurin
p62/SQSTM1	sequestosome 1
PAR-CLIP	Photoactivatable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation
PBS	phosphate buffered saline

PCR	polymerase chain reaction
PDE8A	phosphodiesterase 8A
POU4F1	POU domain, class 4, transcription factor 1
pre-miRNA	precursor-microRNA
pre-mRNA	precursor-messenger RNA
pri-miRNA	primary-microRNA
PSF	polypyrimidine-tract-binding-associated factor
PSI	percent spliced in
pSILAC	pulsed SILAC
PTB	polypyrimidine tract binding protein
PTP4A1	protein tyrosine phosphatase type IVA, member 1
qPCR	quantitative polymerase chain reaction
qRT-PCR	quantitative real-time polymerase chain reaction
RanBP2	RAN binding protein 2
RBP	RNA-binding protein
RefSeq	reference sequence
RGG domain	arginine-glycine-glycine-rich domain
RNA	ribonucleic acid
RNA Pol II	RNA polymerase II
RNAi	RNA interference
RNase	Ribonuclease
RNA-Seq	RNA sequencing
RNP	ribonucleoprotein particle
RRM	RNA-recognition motif

RT	room temperature
SC35	splicing component, 35 kDa
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SILAC	stable isotope labeling with amino acids in cell culture
siRNA	small interfering RNA
snRNA	small nuclear RNA
SOD1	superoxide dismutase 1
SR proteins	serine/arginine-rich proteins
SRM160	serine/arginine-related nuclear matrix protein of 160 kDa
SRSF3	serine/arginine-rich splicing factor 3
ss	single-stranded
STK38	serine/threonine kinase 38
SYGQ-domain	serine-tyrosine-glycine-glutamine-rich domain
T4 PNK	T4 Polynucleotide kinase
TAF(II)s	TBP-associated factors
TAF15	TATA box binding protein-associated factor 15
TARDBP	TAR DNA binding protein
TASR	TLS-associated serine-arginine protein
TBP	TATA box binding protein
TBST	Tris buffered saline with Tween
TDP-43	TAR DNA-binding protein 43
TFIID	transcription factor II D
THAP6	THAP domain containing 6
TLS	translocated in liposarcoma protein

TO	tetracycline operator
TSS	transcription start site
U	uridine
UBQLN2	ubiquilin 2
UPS	ubiquitin–proteasome system
UTR	untranslated region
VCP	valosin-containing protein
YBX1	Y box binding protein 1
ZF	zinc finger
ZNF384	zinc finger protein 384

Acknowledgements

First of all I would like to thank my supervisor Dr. Markus Landthaler for giving me the great opportunity to work on this interesting project and for his help and scientific support during the last years. I also would like to thank my supervisor at Humboldt University, Prof. Dr. Christian Schmitz-Linneweber, for kindly taking supervision and examination of this thesis. Furthermore, I would like to thank Prof. Dr. Oliver Daumke, Prof. Dr. Harald Saumweber and Prof. Dr. Uwe Ohler for taking the time to review my thesis.

I thank all lab member of the Landthaler lab for always helping, supporting and encouraging. I am much obliged to Marvin Jens who did all the computational analysis of the tons of high-throughput data produced in this study. I also thank Dr. Wei Chen, Claudia Langnick and Mirjam Feldkamp for help with the sequencing. Further, I thank Dr. Stefan Kempa and Dr. Guido Mastrobuoni for performing the proteomic analyses.

Last but not least I would like to thank my family and friends for their continuous help and support especially during the exciting last months. My biggest thanks goes to Steve who went with me through all the ups and downs and sleepless nights giving my always unfailing support. Without you I would not manage to finish. Thank you!

Publications

Baethge, K., Jens, M., Maaskola, J., Mackowiak, S., Doormann, D., Dieterich, C., and Landthaler, M. (in preparation) Transcriptome-wide Analysis of Interactions of the RNA-Binding Proteins FUS, EWSR1, and TAF15

Conference contributions

Baethge, K., Mastrobuoni, G., Maaskola, J., Chen, W., Rajewsky, N.,Kempa, S., and Landthaler, M.: Functional characterization of Ewing Sarcoma Breakpoint Region 1 protein, *EMBL Conference - The Complex Life of mRNA: From Synthesis to Decay*, 18.03.-20.03.2010, Heidelberg (poster)

Baethge, K., Mastrobuoni, G., Maaskola, J., Chen, W., Rajewsky, N.,Kempa, S., and Landthaler, M.: Functional characterization of Ewing Sarcoma Breakpoint Region 1 protein, *MDC FMP PhD Symposium 2010*, 27.05.2011, Berlin (talk)

Baethge, K., Mastrobuoni, G., Maaskola, J., Chen, W., Rajewsky, N.,Kempa, S., and Landthaler, M.: Functional characterization of the RNA-binding proteins FUS/TLS and TDP-43, *Berlin Neuroscience Forum 2010*, 10.06.-11.06.2010, Liebenwalde (poster)

Baethge, K., Mastrobuoni, G., Jens, M., Mackowiak, S., Maaskola, J., Rajewsky, N.,Kempa, S., and Landthaler, M.: Functional characterization of the FET family of RNA binding proteins, *The Sixteenth Annual Meeting of the RNA Society (RNA2011)*, 14.06.-18.06.2011, Kyoto (poster)

Supplementary data

Supplementary table 1: Small RNA-seq reads upon FUS, EWSR1, TAF15 and Triple knockdown normalized to the number of total reads and relative to the read number of mock transfection ranked by the number of total reads.

		Normalized small RNA-seq reads upon knockdown relative to mock						
miRNA	**miRNA precursor**	**FUS siRNA1**	**FUS siRNA2**	**EWSR1 siRNA1**	**EWSR1 siRNA2**	**TAF15 siRNA1**	**TAF15 siRNA2**	**Triple KD**
hsa-miR-103a	hsa-mir-103a-2	1,34	1,31	0,66	0,82	0,08	0,88	0,92
hsa-miR-103a	hsa-mir-103a-1	1,34	1,31	0,66	0,82	0,08	0,88	0,92
hsa-miR-16	hsa-mir-16-1	0,58	0,64	1,86	0,57	0,43	0,77	1,19
hsa-miR-16	hsa-mir-16-2	0,58	0,64	1,86	0,57	0,43	0,76	1,19
hsa-let-7a	hsa-let-7a-2	1,05	1,68	0,95	1,97	1,52	1,63	0,75
hsa-let-7a	hsa-let-7a-3	1,05	1,68	0,95	1,98	1,50	1,64	0,75
hsa-let-7a	hsa-let-7a-1	1,05	1,68	0,95	1,98	1,50	1,64	0,75
hsa-miR-92a	hsa-mir-92a-1	1,67	1,62	1,45	2,57	7,46	1,75	0,54
hsa-miR-17	hsa-mir-17	0,92	1,80	1,24	2,39	0,52	2,38	0,53
hsa-miR-92a	hsa-mir-92a-2	1,71	1,65	1,45	2,63	7,19	1,75	0,53
hsa-miR-320a	hsa-mir-320a	1,60	0,89	0,52	0,89	0,66	0,95	2,28
hsa-miR-148a	hsa-mir-148a	2,63	1,66	1,18	1,27	0,42	1,43	0,71
hsa-miR-7	hsa-mir-7-1	0,30	0,23	0,32	0,29	0,13	0,36	1,92
hsa-miR-26a	hsa-mir-26a-2	0,91	1,08	1,04	0,89	0,55	0,85	1,55
hsa-miR-26a	hsa-mir-26a-1	0,91	1,08	1,04	0,89	0,55	0,85	1,55
hsa-miR-7	hsa-mir-7-2	0,29	0,22	0,31	0,29	0,11	0,36	1,98
hsa-miR-7	hsa-mir-7-3	0,29	0,22	0,31	0,29	0,11	0,36	1,98
hsa-miR-10a	hsa-mir-10a	1,08	0,95	1,92	0,88	1,36	1,49	1,19
hsa-miR-93	hsa-mir-93	1,10	1,03	1,66	1,64	0,41	1,02	0,56
hsa-let-7f	hsa-let-7f-2	0,53	0,87	0,29	1,04	0,43	0,73	1,22
hsa-let-7f	hsa-let-7f-1	0,53	0,87	0,29	1,05	0,42	0,71	1,24
hsa-miR-186	hsa-mir-186	3,28	1,70	1,22	1,78	2,58	1,47	0,46
hsa-miR-30e	hsa-mir-30e	1,83	1,61	0,38	1,34	0,38	1,55	1,16
hsa-miR-378	hsa-mir-378	4,33	1,85	1,38	1,78	0,97	1,64	0,48
hsa-miR-15a	hsa-mir-15a	0,31	0,48	4,32	0,83	0,34	0,62	0,42
hsa-miR-26b	hsa-mir-26b	1,60	1,40	1,26	1,28	0,53	1,16	0,62
hsa-miR-10b	hsa-mir-10b	2,43	1,73	1,26	1,20	6,42	1,72	0,53
hsa-miR-148b	hsa-mir-148b	2,50	1,36	0,52	1,21	0,51	0,99	0,73
hsa-miR-15b	hsa-mir-15b	0,31	0,54	0,59	0,69	0,26	0,79	1,61

Eidesstattliche Erklärung

Hiermit versichere ich, Kerstin Baethge, die vorliegende Arbeit selbstständig und ohne unerlaubte Hilfe angefertigt zu haben und alle verwendeten Hilfsmittel und Quellen als solche ausgewiesen zu haben.
Des Weiteren versichere ich, dass die vorliegende Arbeit nie in dieser oder anderer Form Gegenstand eines früheren Promotionsverfahrens war.

Berlin,

Kerstin Baethge